

PCT
 WELTORGANISATION FÜR GEISTIGES EIGENTUM
 Internationales Büro
 INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
 INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



(51) Internationale Patentklassifikation 6 : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027	A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998
		(43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)

(21) Internationales Aktenzeichen: PCT/EP98/08216

(22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)

(30) Prioritätsdaten:
197 57 984.1 24. Dezember 1997 (24.12.97) DE

(71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).

(74) Gemeinsamer Vertreter: BAYER AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).

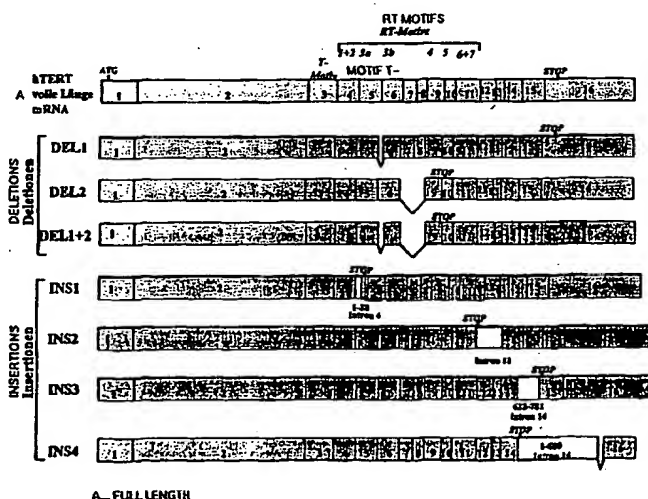
(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht

Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF

(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der

30 Okazaki Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 10
1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

15

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994). 25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet

26 ... ist im Zellkern lokalisiert. Während der RNA-

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumor-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomerasequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomerasequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.
15

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzwelligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebes- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-
30

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

30

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

15 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

20 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die
10 Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.
15

20 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
25 14: nicht verdaut.

30 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

30

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

30

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

15

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

30

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in

10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence		
Intron	Exon	Exon No.	bp	Intron	Intr on
					bp
S' flanking Region					
caggcgcttcccccgag	GTTCAGGCAGCGTGCCT	1	281	CGCCCCCTCCTTCGCCAG	1
catgtccttctcgtttaag	GTGTCTGCTGCTGAAGAGC	2	1354	TGGCTGCGCAGGAGCCCCAG	2
gaggggtctctctattgag	GGTTGGCTGTGTTCCGGC	3	196	TGCAAAAGCATTTGGAATCAG	3
cccatgtgtcccccgag	ACAGCACTTGAAGAGGGTG	4	181	GTTCGCGCAGAGAAAAAGAGG	4
ctgcctccactcacagag	GCCGAGCGTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	5
ccctctcctctgcgggag	GTGGATGTGACGGGCGCGT	6	156	CAAGGCTTCAAGAGCCAC	6
ctcccgctgtcttttcgag	GTCTCTACCTTGACAGACC	7	96	TGCCGTCTGTCATCGAGCAG	7
ctgtgtcttccccgccccag	AGTCCTCCTCCCTGAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	8
gtattttcccttatttttag	GTCCTACGTCCAGTGCCAG	9	114	CGGGGATTGCGCGGGACGG	9
cattgccccctctgccttag	GCTGCTCCTGCGTGTGGTG	10	72	ACGCGAAAACTTCCTCAG	10
attccccctgtgtctctag	GACCTGCTCCGAGGTGTC	11	189	TGCAGAGCGACTACTCCAG	11
tctttcttggcgactcttag	CTATGCCCGGACCTCCATC	12	127	CCTGTTTCTGGATTTCAG	12
ctgtccgccatcctctctag	GTGAACAGCCTCCAGACGG	13	62	TCCTGCTGCAGGCGTACAG	13
agcctctgttttcccccgag	GTTCACGCATGTGTCTG	14	125	CTGAAAGCCCAAGAACGCAG	14
tctgattttggcccccgag	GGATGTGCTGGGGGCCAA	15	138	CTGGGTCTACTCAGGACAG	15
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTGTGAAAAAA	16
					3' flanking Region

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

15 ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20 kloniert.

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses
Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

5	ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70
	ATGAGACCCT	GTCTCAAAAA	AAAAAAGGAA	AATTGAAATA	ATATAAAGCA	TCTTCTCTGG	CCACAGTGGA	140
	ACAAAACCCAG	AAATCAACAA	CAAGAGGAAT	TTTGAAAAC	ATACAAACAC	ATGAAAAATTA	AACAATATAC	210
	TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TAAAAAGGA	AATTGAAAAA	TTTATTTAAG	CAAATGATAA	280
	CGGAAACATA	ACCTCTCAAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	350
	AGCAGCTACA	TCAAAAAAGT	AGAAAAGCCA	GGCGCAGTGG	CTCATGCCCTG	TAATCCACAG	ACCTTTGGGAG	420
10	GCCAAGGCGG	GCAGATCGCC	TGAGGTCAGG	AGTTCGAGAC	CAGCCTGACC	AACACAGAGA	AACCTTGTCG	490
	CTACTAAAAA	TACAAAATTA	GCTGGGCATG	GTGGGCATG	CCTGTAATCC	CAGCTACTCG	GGAGGCTGAG	560
	GCAGGATAAC	CGCTTGAACC	CAGGAGGTGG	AGGTTGCGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	630
	TGGGTAACAA	GAGTGAAACC	CTGTCTCAAG	AAAAAAGAAA	AAGTAGAAAA	ACTTAAAAAT	ACAACCTAAT	700
	GATGCACTT	AAAGAACTAG	AAAAGCAAGA	GCAAACTAAA	CCTAAAATTG	GTAAGAAGAA	AGAAATAATA	770
15	AAGATCAGAG	CAGAAATAAA	TGAACTGAA	AGATAACAAT	ACAAAAGATC	AACAAAATTA	AAAGTTGGTT	840
	TTTTGAAAG	ATAAACAAAA	TTGACAAACC	TTTGCCAGAG	CTAAGAAAAA	AGGAAAGAAG	ACCTAAATAA	910
	ATAAAGTCAG	AGATGAAAAA	AGAGACATTA	CAACTGATAC	CACAGAAATT	CAAGGATCA	CTAGAGGCTA	980
	CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACCTAGAA	AAAAATAGATA	AATTCTTAGA	TGCATACAAC	1050
	CTACCAAGAT	TGAACCATGA	AGAAATCCAA	AGCCCAAAAC	GACCAATAAC	AATAATGGGA	TTAAAGCCAT	1120
20	AATAAAAGT	CTCCTAGCAA	AGAGAAGCCC	AGGACCCAAT	GGCTTCCCTG	CTGGATTTTA	CCAATCATTT	1190
	AAAGAAGAAT	GAATTCCAAT	CCTACTCAAA	CTATTCTGAA	AAATAGAGGA	AAGAATACTT	CCAACTCAT	1260
	TTACATGGC	CAGTATTACC	CTGATTCCAA	AACCAAGACA	AAACACATCA	AAAACAAACA	ACCAAAAAAA	1330
	CAGAAAGAAA	GAAACTACA	GGCCAATATC	CCTGATGAAT	ACTGATACAA	AAATCCTCAA	CAAAACACTA	1400
	GCAAAACCAA	TTAAACAACA	CCTTCGAAAG	ATCATTCAAT	GTGATCAAGT	GGGATTATT	CCAGGGATGG	1470
25	AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAATGAAGT	ACAAAAACTA	1540
	TATGATTATT	TCACTTTATG	CAGAAAAAGC	ATTTGATAAA	ATTCTGCACC	CTTCATGATA	AAAAACCTCA	1610
	AAAAACCAGG	TATACAAGAA	ACATACAGGC	CAGGCACAGT	GGCTCACACC	TGCGATCCCA	GCACCTGGG	1680
	AGGCCAAGGT	GGGATGATTG	CTTGGGCCCC	GGAGTTTGAG	ACTAGCCTGG	GCAACAAAAAT	GAGACCTGGT	1750
	CTACAAAAAA	CTTTTTTAAA	AAATTAGCCA	GGCATGATGG	CATATGCCTG	TAGTCCACAG	TAGCTCTGGG	1820
30	GCTGAGGTGG	GAGAATCACT	TAAGCCTAGG	AGGTCGAGGC	TGCAGTGAGC	CATGAACATG	TCACGTACT	1890
	CCAGCCTAGA	CAACAGAACA	AGACCCCACT	GAATAAGAAG	AAGGAGAAGG	AGAAGGGAGA	AGGGAGGGAG	1960
	AAGGAGGAG	GAGGAGAAGG	AGGAGGTGGA	GGAGAAGTGG	AAGGGGAAGG	GGAAGGGAAA	GAGGAAGAAG	2030
	AAGAAACATA	TTTCAACATA	ATAAAGCCCC	TATATGACAG	ACCGAGGTAG	TATTATGAGG	AAAAACTGAA	2100
	AGCCTTTCTC	CTAAGATCTG	GAAATGACA	AGGGCCCACT	TTCAACCATG	TGATTCAACA	TAGTACTAGA	2170
35	AGTCCTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAAA	GGCATCCCAA	CTGGAAGGA	AGAAGTCAAA	2240
	TTATCCTGTT	TGCAGATGAT	ATGATCTTAT	ATCTGGAAAA	GACTTAAGAC	ACCACTAAAA	AACATTATGA	2310
	GCTGAAATTT	GGTACAGCAG	GATACAAAAAT	CAATGTACAA	AAATCAGTAG	TATTTCTATA	TTCCAACAGC	2380
	AAACAACTG	AAAAAGAAAC	CAAAAAAGCA	GCTAGATAAA	AAATTAACA	GCTAGGAATT	AACCAAGAAA	2450
	GTGAAAGATC	TCTACAATGA	AACTATAAAA	ATGTTGATAA	AAGAAATTGA	AGAGGGCACA	AAAAAAGAAA	2520
40	AGATATTCCA	TGTTTCATAGA	TTGGGAAGAT	AAATCTGTT	AAATGTCCA	TACTACCCAA	AGCAATTTAC	2590
	AAATTCAATG	CAATCCCTAT	TAAATACTA	ATAGCCTTCT	TCACAGAAAT	AGAAGAAACA	ATTCTAAGAT	2660
	TTGTACAGAA	CCACAAAAGA	CCCAGAATAG	CCAAAGCTAT	CCTGACCAAA	AAGAACAATA	CTGGAAGCAT	2730
	CACATTACCT	GACTTCAAAAT	TATACTACAA	AGCTATAGTA	ACCCAAACTA	CATGGTACTG	GCATAAAAC	2800
	AGATGAGACA	TGGACCAGAG	GAACAGAATA	GAGAATCCAG	AAACAAATCC	ATGCATCTAC	AGTGAACCTA	2870
45	TTTTTGACAA	AGGTGCCAAG	AACATACTTT	GGGGAAGAGA	TAACTCTCTC	AATAAATGGT	GCTGGAGGAA	2940
	CTGGATATCC	ATATGCAAAA	TAACAATACT	AGAATCTGT	CTCTCACCAT	ATACAAAAGC	AAATCAAAAT	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAAA	CTTTGCAACT	ACTAAAAAGAA	AACACCCGAG	AAACTCTCCA	3080
	GGACATTGGA	TGGGCAAAAG	ACTTCTTGAG	TAATTCCTGT	CAGGCACAGG	CAACCAAGAG	AAAAACAGAC	3150
	AAATGGGATC	ATATCAAGTT	AAAAAGCTTC	TGCCCAGCAA	AGGAAACAAT	CAACAAAGAG	AAGAGACAAC	3220
50	CCACAGAATG	GGAGAATATA	TTTGCAAACT	ATTACTCTAA	CAAGGAATTA	ATAACCAAGT	TATATAAGGA	3290
	GCTCAAACTA	CTCTATAAGA	AAAACACATA	ATAAGCTGAT	TTTCAAAAT	AAGCAAAAGA	TCTGGGTAGA	3360
	CATTTCTCAA	AATAAGTCAT	ACAAATGGCA	AACAGGCATC	TGAAAATGTG	CTCAACACCA	CTGATCATCA	3430
	GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCAGTTTAAA	ATGGCTTTTA	TTCAAAAGAC	3500
	AGGCAATAAC	AAATGCCAGT	GAGGATGTGG	ATAAAAGGAA	ACCCCTGGAC	ACTGTTGGTG	GGAATGGAAG	3570
55	TTGTACCAC	TATGGAGAAC	AGTTTGAAGG	TTCTCAAAA	AACTAAAAAT	AAAGCTACCA	TACAGCAATC	3640
	CCATTGCTAG	GTATATACTC	CAAAAAAGGG	AATCAGTGTA	TCAACAAGCT	ATCTCCACTC	CCACATTTAC	3710
	TGCAGCACTG	TTCATAGCAG	CCAAGGTTTG	GAAGCAACCT	CAGTGCTCAT	CAACAGACGA	ATGAAAAAAG	3780
	AAAATGTGGT	GCACATACAC	AATGGAGTAC	TACGCAGCCA	TAAAAAAGAA	TGAGATCCTG	TCAGTTGCAA	3850
	CAGCATGGGG	GGCACTGGTC	AGTATGTTAA	GTGAAATAAG	CCAGGCACAG	AAAGACAAAC	TTTTCATGTT	3920
60	CTCCCTTACT	TGTGGGAGCA	AAAATTAATA	CAATTGACAT	AGAAATAGAG	GAGAATGGTG	GTTCTAGAGG	3990
	GGTGGGGGAC	AGGGTGACTA	GAGTCAACAA	TAATTTATTG	TATGTTTTAA	AATAACTAAA	AGAGTATAAT	4060
	TGGGTTGTTT	GTAACACAAA	GAAAGGATAA	ATGCTTGAAG	GTGACAGATA	CCCCATTTAC	CCTGATGTGA	4130
	TTATTACACA	TTGTATGCCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCCCTA	CTATATTAAA	4200
	AATTAAAAAT	TTAATGGCCA	GGCACGGTGG	CTCATGTCCG	TAATCCACAG	ACTTTGGGAG	GCCGAGGCGG	4270
65	GTGGATCACC	TGAGGTCAGG	AGTTTGAAGC	CAGCTGTGCC	ACCATGATGA	AACCCTGTCT	CTACTAAAGA	4340
	TACAAAAAAT	AGCCAGGCGT	GGTGGCACAT	ACCTGTAGTC	CCAACTACTC	AGGAGGCTGA	GACAGGAGAA	4410
	TTGCTTGAAC	CTGGGAGGCG	GAGGTTGCAG	TGAGCCGAGA	TCATGCCACT	GCACCTGCAG	CTGGGTGACA	4480
	GAGCAAGACT	CCATCTCAAA	ACAAAAACAA	AAAAAAGAG	ATTAAAAATTG	TAATTTTTAT	GTACCGTATA	4550
	AATATATACT	CTACTATATT	AGAAGTTAAA	AATTAACA	ATTATAAAG	GTAATTAACC	ACTTAATCTA	4620
70	AAATAAGAAC	AATGTATGTG	GGGTTTCTAG	CTTCTGAAGA	AGTAAAAAGTT	ATGGCCACGA	TGGCAGAAAT	4690
	GTGAGGAGGG	AACAGTGGAA	GTTACTGTTG	TAGACGCTC	ATACTCTCTG	TAAGTGACTT	AATTTTAACC	4760
	AAAGACAGGC	TGGGAGAAGT	TAAAGAGGCA	TTCTATAAGC	CCTAAAACAA	CTGCTAATAA	TGGTGAAAGG	4830
	TAATCTCTAT	TAATTACCAA	TAATTACAGA	TATCTCTAAA	ATCGAGCTGC	AGAATTGGCA	CGTCTGATCA	4900
	CACCGTCTC	TCATTACCG	TCGTTTTTTT	CTTGTGTGCT	TGGAGATTTT	CGATTGTGTG	TTCTGTGTTG	4970
75	GTAAACTTA	ATCTGTATGA	ATCCTGAAAC	GAAAAATGGT	GGTGATTTC	TCCAGAAAGAA	TTAGAGTACC	5040
	TGGCAGGAAG	CAGGTGGCTC	TGTGGACCTG	AGCCACTTCA	ATCTTCAAGG	GTCTCTGGCC	AAGACCCAGG	5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCTGTC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
5	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTCACTTGA	5390
	AGCAGTGACC	AAGGTGTATT	CTGAGGGAAG	CTTGAGTTAG	GTGCCTTCTT	TAAACAGAAA	AGTCATGGAA	5460
	GCACCCCTCT	CAAGGGAAAA	CCAGACGCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTTG	5530
	CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCCGTGGG	GCTTCTCCGA	GCCCCGTGCTG	AGGACCTCT	5600
	TGCAAAGGGC	TCCACAGACC	CCCGCCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACTGGGATG	5670
	TGGCTGGGGG	CGGACAGCGA	CGGCGGGATT	CAAAGACTTA	ATTCCATGAG	TAAATTCAAC	CTTTCCACAT	5740
10	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTCCAGGA	CTGCAGAAAT	5810
	CCAAAGGCGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAGG	TCCAAGGACT	TAATAACCAT	GTTACAGAGG	5880
	ATTTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTTATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950
	AGAGGCCTGG	GCGGCAGGGC	TATGAGCAGC	GCAGGGCCAC	CGGGGAGAGA	GTCCTCCGCC	TGGGAGGCTG	6020
15	ACAGCAGGAC	CACTGACCGT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	CGAAGGCGGC	CACGCTGCGT	6090
	GTGACTCAGG	ACCCCATACC	GGCTTCTCTG	GCCCCACCCAC	ACTAACCAGG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAC	GAACATGACC	CTTGCTGGC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	TGGGTGTGCA	6230
	GGAATAGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCTTTC	CATCATTATT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	TACAAACACC	6370
20	ACTCTTTTAC	TAGGCCACA	GAGCAGGSC	CACACCCCTG	ATATATTAG	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCCTGG	6510
	CAGGCACTCC	CCAGATTCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTC	CCGCCCTCAG	GCCTCAGCTT	CTCCAGCAGC	6650
	TTCTTAAACC	TGGGGTGGGC	CGTGTTCAG	GCGTACTGTC	TCACCTGTCC	CACCTGTGCT	TGCTCAGCGT	6720
25	ACGTAGCTCG	CACGGTTTCT	CCTCACATGG	GGTGCTGTCT	TCTTCCCTCA	ACACTCACAT	GCGTTGAAGG	6790
	GAGGAGATTG	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GATGCAGGTT	6860
	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCGTCTCT	CTGTCACTCG	CCGGGCGCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	CTAGGGTCTC	7000
	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	CATTGGGGTG	7070
30	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	CCGCCAGGCA	7140
	CCCGCCCTTC	TCTGCCCAGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TCCACAAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CCTGGGAATT	7280
	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGG	CTGTTTTATT	TAAATAGCTA	7350
	CAAAGCAGGG	AAATCCCTGC	TAAAAATGTC	TTTAAACAAC	TGGTTAAACA	AACGGGTCCA	TCCGCACGGT	7420
35	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TAGCAACAT	GCTCAAAAG	AAAGAAATTC	ACCCCATGGC	7560
	AGGGAGTGGT	TAGGGGGGTG	TAAGGACGGT	GGGGGGCGCA	GCTGGGGGCT	ACTGCACGCA	CTTTTACTA	7630
	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700
	GGAACCCGGA	GGCTGTGCCA	TCTTTGCCAT	CCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCCC	7770
40	ACGTCCTGAT	TCCCCAAAC	CTGTGGACAG	AAACCGCCCG	GCCCCAGGGC	CTTTGCAGGT	GTGATCTCCG	7840
	TGAGGACCCT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTCTGGGA	7910
	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCGCTCAGGAC	GATGGAGGCA	GTCAGTCTGA	GGGTGAAAG	7980
	GGAGGGAGGG	CTCGAGCCC	AGGCCGTGCA	GCGCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCTCC	8050
	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCCTCC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTGCAGAA	GCAACAGGAA	8190
	ACCCATCTGA	TGTGAATCTA	GGATTATTTT	AAAACAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260
	GTGCCCTCGG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TACGCCCTCC	AAGTAGCTGG	GATTTTCAGG	GTGCACCACC	8470
50	ACACCCGGCT	AATTTTGTAT	TTTATAGTAG	GATGGGCTTT	CACCATGTTG	GTCAGCTGTA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGCACCT	8610
	GGCCTATTTA	ACCATTTTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTC	8680
	AATTTCCCTG	TACTCAGGA	GTTACCTCTC	TTTGTATTTT	TCTGTAATTC	TTCGTAGACT	GGGGATACAT	8750
	CGTCTCTTGA	CATATTCACA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	GGGCACTGG	8820
	GAGGCTGAGC	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGGCC	8890
55	AAGTGTGGAC	ACTGTCCCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	TAGAAATTA	AGATCCATCC	8960
	CTCCTACTCT	ACTGGGATTG	AGCCCCCTTC	CTATCCCCCC	CCAGGGGCAG	AGGAGTTTCT	CTCACTCTCT	9030
	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACTGCTGGT	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
60	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGATTAG	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GGGGGTGGGT	GGGGTTCACC	9310
	ATGTTGGCCA	GGCTGGTCTC	GAACTTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCCTCT	AAAGTGCTGG	9380
	GATTACAGGT	GTAGGCCACC	ATGCCACAGT	CAGCAATTTAC	TCTGTTTAGA	AACATCTGGG	TCTGAGGTAG	9450
	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAAGCAA	TGATAGAATT	TTTTTATTGT	TGTTAGAACA	9520
65	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCCACAT	9660
	GGTGTAAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTCTTCCCT	CTTTTAAAT	TGTGTTTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACCACTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GCCCTTTTGC	CTTAGTGCCA	GAGACAATTC	ACAAACACAG	CCCTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTTAG	AAGAGCGACC	TGTAATCCTA	AGATTTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	GCCAGGGAGG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	TTTGTGTAGC	ATTTTCAGTG	TTGCCGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTTCTGCCC	CCTTAGATCC	AAACTTGAGC	AACCTCGAGT	10150
	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGGCCCA	GGTCTGGAGG	GGACCACTGG	10220
75	CCGTGTGGCT	TCTACTGTCT	GGCTGGGAAT	CGGGCTCTCT	AGCTCTGCAG	TCCGAGGCTT	GAGGCCAGGT	10290
	GCCTGGACCC	CGAGGCTGCC	CTCCACCCTG	TCCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	CATCTGCCAC	10360
	ACAGAGTGCC	GGGGCCAGG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGCCTG	GCCAGCAGGA	10430
	GCGCTGGCT	CCATTTCCCA	CCCTTTCTCG	ACGGGACCGC	CCCGGTGGGT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCTCG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCTGTCAG	GGAGCAATGC	10640
	GTCTCGGGT	TCGTCCCCAG	CCGCGTCTAC	GCCTCCCGT	CCTCCCTTC	ACGTCCGGCA	TTCGTGGTGC	10710
5	CCGGAGCCCG	ACGCCCCGCG	TCCGGACCTG	GAGGCAGCCG	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAAG	10780
	GGTCGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCTCCCT	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTGCA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCGCGA	GCGGCGCGCG	10920
	GGCGGGGAAG	CGCGGCCAG	ACCCCGGGT	CCGCCGGAG	CAGCTGCGCT	GTGCGGCCA	GGCCGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCAGG	ACCGCGCTCC	CCAGTGGCG	GAGGGACTGG	GGACCCGGGC	11060
	ACCCGTCTG	CCCCTTACCC	TTCCAGCTCC	GCCTCTCCG	CGCGGACCCC	GCCTGTCCTC	GACCCCTCCC	11130
10	GGGTCCCGG	CCCAGCCCC	TCCGGGCCCT	CCCAGCCCT	CCCCTTCTT	TCCGCGGCC	CGCCCTCTCC	11200
	TGCGCGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATCCCG	GCGCTCCCG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCCG	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCTGTGCG	CGCTGGGGC	CCCAGGGCTG	GCGGTGCTG	CAGCGCGGGG	ACCCGGCGGC	11410
15	TTTCCGCGCG	CTGGTGGCCC	AGTGCTGGT	GTGCTGCCCC	TGGGACGCAC	GGCGGCCCCC	CGCCGCCCCC	11480
	TCCTTCCGCC	AGGTGGGCT	CCCCGGGGT	GGCGTCCGGC	TGGGGTTGAG	GGCGGCCGGG	GGGAACCCAG	11550
	GACATGCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGTTTCCC	CCGACGGTGT	CCTGCTGAA	GGAGCTGGTG	11620
	GCCCCAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGGGAAGA	ACGTGCTGGC	CTTCGGCTTC	GCGCTGTGG	11690
	ACGGGGCCCG	CGGGGGCCCC	CCCCGAGCCT	TCACCACCAG	CGTGCAGCAG	TACCTGCCCA	ACACGGTGAC	11760
20	CGACGCACTG	CGGGGAGCG	GGGCGTGGG	GCTGCTGCTG	CGCCGCGTGG	GCGACGACGT	GCTGGTTCAC	11830
	CTGCTGGCAG	CTTGCGCGCT	CTTTGTGCTG	GTGGCTCCCA	GCTGCGCCTA	CCAGGTGTGC	GGGCGCCCGC	11900
	TGTACAGACT	CGGCGCTGCC	ACTCAGGCC	GGCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGTCTGGG	11970
	ATGCGAACGG	GCCTGGAACC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCGGTGGCG	12040
	AGGAGCGCGG	GGGGCAGTGC	CAGCCGAAGT	CTGCCGTGG	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTTG	12110
25	AGCCGGAGCG	GACGCCGTT	GGGAGGGGT	CCTGGGCCCA	CCCGGGCAGG	ACGCGTGGAC	CGAGTGACCG	12180
	TGGTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	CTCTCTGGC	12250
	ACCGCCACT	CCACCCATC	CGTGGGCCG	CAGCAGCAGC	CAGGCCCCCC	ATCCACATCG	GCGCCACCAC	12320
	GTCCCTGGGA	CACGCCTTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCCT	CAGGCGACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCACGCCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGCG	GTGCCCCCG	CTGCCCCAGC	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGTGCCTTGG	GAACCAACGC	CAGTGCCCTT	ACGGGTGCTG	12600
	CCTCAAGACG	CACGTGCCCG	TGCGAGCTGC	GGTCACCCCA	GCAGCCGGTG	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGCCCG	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GSTGCAGCTG	CTCCGCCAGC	12740
	ACAGACCCCG	CTGGCAGGTG	TACGGCTTCG	TGCGGCCCTG	CCTGCCCGCG	CTGGTGCCCC	CAGGCTCTGT	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCCT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAGCATGCC	12880
	AAGCTCTCGC	TGCAGGAGCT	GACGTGGAAG	ATGAGCGTGC	GGGACTGCGC	TTGGCTGCGC	AGGAGCCAG	12950
	GTGAGGAGCT	GGTGCCCGT	GAGGGGCCAG	CCCCAGAGC	TGAATGACGT	AGGGGCTCAG	AAAAGGGGCG	13020
	AGGCAGAGCC	CTGGTCTCTC	TGTCTCCATC	GTCACGTGGG	CACACGTGGC	TTTTCGCTCA	GGACGTGCG	13090
	TGGACACCGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGATATA	ACTTACGAGG	TTACCTTCA	13160
40	CGTTTGTATG	GACACGCGGT	TTCCAGGCCG	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GCGCGCGGCG	13230
	TGGAGCCGGG	TTGCCGGCAA	TGGGGAGAAG	TGTCTGGAAG	CACAGACGCT	CTGGCGAGGG	TGCCTGCAGG	13300
	TTACCTATAA	TCCTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTTCTCTG	13370
	GGGTGGGAGG	TAAGGTTT	GCAGGTGCAC	TGTGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTAAAT	13440
45	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
	TCACCTGAGG	TCAGGAGTTT	GAGACCAGCC	TGACCAACAT	GGTGAACCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTGG	TGTGTGCTCT	TAATCCCAGC	TACTTGGGAG	GCTGAGGCG	GAGAATCACT	13650
	TGAACCCAGG	AGGCGGAGGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	CGACAAGACT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGTT	CGTTGATTGT	GCCAGGACAG	GSTAGAGGGA	GGGAGATAAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATAGCTCC	ACCTGCTGAG	GAAGGGACAG	TGTTGTGGG	TGTTACAGGG	ATGGTGCTGC	TGGCGCCCTG	13930
	CGTGTCCTCA	CCCTGTTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCAG	14000
	TGCTCCACAG	CCCTACCGTG	GCAGTAGAA	GAAGTCCCGA	TTTACCCCC	TCCCCACAAA	CTCCCAAGC	14070
	ATGTAAGTCT	TCCGCGCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGCTC	TTTTTTTCTT	TTTTTCTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTG	CTTGAATATG	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCACGCGC	AGTCAGATAA	GCGTCATGCA	14350
	ACCCAGTTTT	GCTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTCTCTGTAT	CAGGACTCTG	14420
	CCTGTCTATT	CTGTCTCTG	ACTTCAGATG	AGGTCACAAT	CTGCCCTTGG	CTTATGCAAG	GAGTGAGGCG	14490
	TGGTCCCCCG	GTGTCCTGT	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTCACGTGTA	14560
60	GGGTGAGTGA	GGGCGGGCCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGGT	GAGTGAGGCG	14700
	TGGTCCCCCG	GTGTCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGTCCC	GGGTGTCCCT	GTCACGTGCA	14770
	GGGTGAGTGA	GCGCGGTCC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGCG	CCCAGGGTGT	14840
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCAAC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCCG	GTGTCCTGT	CACGTGCAGG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTGCTG	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGCTCC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GCGCCGGTTG	15120
	CCCATGCTCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTCTGTAT	GCTCGCTCT	15190
	TCTTGGTCA	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGACTGCAGG	CTCTCGCCTC	CCCGTGCCCA	15260
70	GGCAGTCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGCTG	GGCCTGGCTT	GCTCACCACG	GCCCGGCCAC	15330
	ATGCATGCTG	CAAATACTCC	TCTCCAGCT	TGTCTCATGC	CGAGGCTGGA	CTCTGGGCTG	CCTGTGCTG	15400
	CTGCCACGTC	TTGCTGGAGA	CATCCACAGAA	AGGGTTCTCT	GTGCCCTGAA	GGAAAGCAAG	GCAACCCAGC	15470
	CCCTCACTT	GTCTGTGTTT	CTCCCAAGCT	GTGGCCCTCT	TGGGCTGGTG	TGGGTGGGTG	TCACACGCTT	15540
	TCACCTTATT	CTGGGCACCT	GCCGCTCATT	GCTTAGGCTG	GGCTCTGCCT	CCAGTCGCCC	CCTCACATGG	15610
	ATTGACGTC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGTC	GAGGGCCGGT	15680
75	GTCTCCGCGA	GCCTTCGTCA	GACTTCCCTC	TTGGGTCTTA	GTTTTGAATT	TCAGTGATT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCTTTTCTT	TTTATTCCTT	TTTATTCCTT	TTTATTCCTT	TTAGTTTAGT	15820
	CATGCCTTTC	CCTCTAAGTG	CTGCCCTTACC	TGCACCTGT	GTTTTGATGT	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCTGTGA	TTTCTTTGAG	CAGTGAGTGA	TTTGAACACT	GTTTATGTTC	AAGATATGTA	GAGTATCAAG	16100
5	ATACGTAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTAAATTTG	ACATCCTGTC	AATAGTGGGC	ATGCATGTTT	ACTATATCCA	GCTTATTAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTCAAT	TCTTCTCGIT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
	TTATCTTCTT	GATGAGTGAA	TCITTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
10	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCCTGC	TGTGTCTGTT	TTCTGCCTTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACITT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
	TGCTGTGTTG	CCCAGGCTGG	TCTCAAATC	TTGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
15	CTGAATTACA	GGCATGAGCC	ACCATGTCG	GCCTAATTTT	CAACACTTTT	ATATTCTTAT	AGTGTGGGTA	16940
	TGTCCTGTTA	ACAGCATGTA	GGTGAATTC	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CCTCGTTCCT	TTGTTTCTCA	CCACCTCTTG	GGTGGCATCT	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCCT	17150
20	CTCGTTGCCT	CCTGGTCACT	GGGCATTTGC	TTTTATTCTT	CTTTGCTTAG	TGTTACCCCT	TGATCTGCTT	17220
	ATTGCTGTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCAACCA	GGCTGGAGTG	TAATGGCACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCCTCCTC	GGTTCAAGCA	GTTCTCATTC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCCACCAC	CACGCCTGGC	TAATTTTTGT	ATTTTTAGTA	GAGATAGGCT	TTCACCATGT	17430
	TGGCCAGGCT	GGTCTCAAAAC	TCCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACACG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGCGATACC	TTGATCTTTT	AAATGAAGT	CTGAAACATT	GCTACCCCTG	17570
	TCCTGAGCAA	TAAGACCCCT	AGTGTATTTT	AGCTCTGGCC	ACCCCCCAGC	CTGTGTGCTG	TTTTCCCTGC	17640
	TGAGTGTAGT	CTATCTCAGG	CATCTTGACA	CCCCCACAAG	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTCTCT	GTAGCTTTGC	CCCCGCCCTG	CTTTCTCTCC	TTTGTTCCTC	GTCTGTCTTC	TGTCTCAGGC	17780
	CCGCCGCTGT	GGGTCCCTCT	CCTTGTCTCT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
30	CTTACCTGTT	GCTGGCTCTC	ATGGCATCTA	GCGACGTCCG	GGGACCTCTG	CTTATGATGC	ACAGATGAAG	17920
	ATGTGGAGAC	TCACGAGGAG	GGCGGTCTATC	TTGGCCCGTG	AGTGTCTGGA	GCACCACGTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGGCCCTGGT	TCAGCCTGGA	AAACCCACAG	CATGTCCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGAAA	TGCGGCAAAAC	CTGCGGTGTG	GCGCCAGCTC	TGACGGTGCT	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCTCCT	TCTGTTTGGG	AACCAGGACA	AAGGATGAGG	CTCCGAGCCG	18200
35	TTGTCCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTAAATTT	TCTAGGCTGG	GCGCGGTGGC	18270
	TCACGCCTGT	AATCCCAGCA	CTTTGGGAGG	CCAAGCGCGG	TGGATCACGA	GGTCAGGAGG	TCCGAGCAT	18340
	CTGGCCCAAT	ATGATGAAAC	CCCATCTGTA	CTAAAAACAC	AAAAATTAGC	TGGGCGTGGT	GGCGGCTGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGCG	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTGCAGTGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAATAAA	AATTCTAGTA	GCCACATTAA	AAAAGTAAAA	AAGAAAAAGT	GAAATTATGT	TAATAATAGA	18620
	TTTTACTGAA	GCCCAGCATG	TCCACACCTC	ATCATTTTAG	GGTGTATTG	GTGGGAGCAT	CACTCACAGG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGCCT	18760
	GGACCTGCTG	GCTTCCCAT	GGCCATGGCT	GTTGATCCAG	ATGGTGCAGG	TCCGGGATGA	GTCGCCCAGG	18830
	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGCACG	TCTGGGATGA	GGTCGCCCAGG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCAGGGGTG	AGGTCTCCAG	GCCTCCGGT	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCAG	GTCCGGGGTG	AGGTCCGCCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGC	AGGTCTGGGG	19180
50	TGAGGTACCC	AGGCCCTGCG	GTGAGCTGGG	TGTGCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTCCG	19250
	CAGACGTGTC	CAGACCATGC	GGTGAGTGGG	ATATGCGGTG	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTTG	19320
	CCAGGCCCTG	CTGTGAGTTG	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAAGTCT	GGGGTGAAGG	TCGCCAGGGC	CCTGCTTTGT	19460
	AGCTGTAGTG	TGTGTGTCTG	GATGGTGCAG	GTCGTGAGTG	AGGTCCGCCAG	GCCCTCGGTG	GCTGGATGTT	19530
	GCAGTGTCCA	GATGGTGCAG	GTCCGGGGTG	AGGTCCGCCAG	ACCCTCGGGT	GAGCTGGATG	TCCGGTGTCT	19600
	GGATGTGTGA	GGTCTGGAGT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGTGTGC	19670
55	GGTCCGGGGT	GAGGTGCGCA	GACCTGTCTG	TGAGCTGGAT	GTGCGGTGTG	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTGCGG	AGACCTGTCT	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTACAGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTCCG	CAGGCCCTGC	19880
	TGTGAATCGG	ATGTGCGCGG	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTGC	CCAGGCCCTG	GGTGAAGTGG	19950
60	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAAGTGC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	20020
	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTC	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGCAAGTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GATGTGCTGT	TATCCGGATG	GTGCAAGTCT	GGGTGAGGTT	20230
	CACCAAGCCC	TGCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAGGTC	CGGGTGAGT	TGCGCAGGCC	20300
	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGGT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
65	TGGTGGGCTG	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	20440
	GATGTGCCGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGTGGT	20510
	GTCCGGATGG	TGCAGGTCCG	GCGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGTCTGGATG	20580
	GTGCAGGTCC	GGGGTGAAGT	AGCCAAGGCC	TTCCGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGCAAGTC	20650
70	CGGGGTGAGG	TGCGCAGGCC	CTGCCAGGCC	CTGGATATGC	GGTGTCCGGA	TGGTGCAGGT	CCGGGGTGAG	20720
	GTCAACAGGC	CCTGCCGTTA	GCTGGATGTG	CGGTGTCTGG	ATGGTGCAGG	TCCGGGGTGA	GGTCCGACAG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGGTGCAG	GATCCGGGGT	AGGTCCGCCAG	GCCCTGCAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTCCGCCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTACCCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
	CGGATGGTGC	AGGTCTGGGG	TGAGGTCCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGGATGGTG	21070
75	CAGGTCCGGG	GTGAGGTCCG	CAGGCCCTGC	GGTGAGCTGG	ATGTGCTGTA	TCCGGATGGT	GCAGGTCTGG	21140
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAGT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGTCTGGATG	GTGCAAGTTC	GGGGTGAAGT	CGCCAGGCCCT	21280

	TGCGGTGAGC	TGGATGTGTG	GTGTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCGCT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGCGGT	21490
5	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTTAAGGGGT	TGGCTGTGTT	CCGGCCGAG	AGCACCGTCT	GCGTGAGGAG	ATCCTGGCCA	AGTTCTCTGA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTGAGGCT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAACAGGC	TCTTTTCTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAA	CATTGGAATC	AGGTACTGTA	21770
	TCCCCACGCC	AGGCCCTCTG	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TCAGCATGCG	CCTGTCTCCA	21840
10	CTTGCCCTGTG	CTTCCCTGGC	TGTGCAGCTC	TGGGCTGGGA	GCCAGGGGCC	CCGTACACAG	CCTGGTCCAA	21910
	GTGGATTCTG	TGCAAGGCTC	TGACTGCCTG	GAGCTCACGT	TCTCTTACTT	GTAATATCAG	GAGTTTGTGC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTTAAA	TTAGATGGAA	ACACTACCAC	TAGCCTCCTT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTGCCCAG	GCTGGAGTGC	AGTGGCATAA	TCTTGGCTCA	CTGCAACCTC	CACCTCCTGG	GTTTAAGCGA	22190
15	TTACACAGCC	TCAGCCCTCT	AAGTAGCTGG	GATTACAGCG	ACCTGCCACC	ACGCTGGGCT	AATTTTGTGA	22260
	CTTTTAGGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGGTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGGC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCCGAT	TCTCTTTTAA	22400
	TTTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGATTTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGACGGGAGC	ACCTGTGCAG	GGAGCAGCTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
20	TCTAGGTGGC	TGCATTTGAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCGG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTC	AAGCTGGATT	TGCATCAGTG	AGGGACGGGA	GCGCTGGTCT	GGGAGATGCC	AGCCTGGCTG	22680
	AGCCCAGGCC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAGC	GCTGACTGTG	GAGGGCTTTA	GTGAGAAGAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CGTGGTGCTG	GGGCCATTTC	CTTGCACTCT	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AGGTAAATAC	22960
	ACAATGCACC	TACTTAGAC	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCAATAT	23030
	TTTGGAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCCAAGA	TGCTCCTTGT	23100
	CACTACTGGG	ACTGTTGTTT	TGCCCTGGGG	GCCTTGAGAG	CCCCTCCTCC	CTGGACAGGG	TACCGTGCTT	23170
	TTTCTACTGT	GCTGGGCCCT	CGGCTGCGG	TGAGGGCACC	AGCTCCGGAG	CACCCGCGCG	CCACGTGCTC	23240
30	ACGGAGTGCC	AGGCTGTGAG	CCACAGATGC	CCAGGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCCCAT	23310
	GGGTGTTTTT	GGGGGAAAAG	GCCAAAGGGA	GAGGTGTGAG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGCC	TGAGCAGCCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGGTCC	TGCCCTGGGG	CAGCCTTGGG	CTACCCAGT	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGCAT	GGGTTACAGT	GGCCCCAGAT	GCAGCCTGGG	ACCAGGCTCC	CTGGTGCTGA	TGGTGGGACA	23590
35	GTACCCCTGG	GGGTTGACCG	CCGAGCTGGG	CGTCCCCAGG	GTTGACTATA	GGACCAGGTG	TCCAGGTGAC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGCGTCT	GCGTGGCAGT	GGTGGACGCT	GCCCCGGGGA	TGGCTTCCCG	23730
	CGTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTGGG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGCTGTGGCT	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAAGAGGGTG	23870
40	CAGTGTGCGG	AGCTGTCCGA	AGCAGAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCGCCCTG	GTGACGTCCA	23940
	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	TGTGAACATG	GACTACGTCG	TGGGAGCCAG	24010
	AACCTTCCGC	AGAGAAAAGA	GGGTGGCTGT	GCTTTGTGTT	AACTTCTCTT	TTAAACAGAA	GTGCGTTTGA	24080
	GGCCCACTT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTGTT	GCGCACAGTG	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
45	GGGTGTAGGG	GGAGCTCCTG	GGGCAGGGAC	AGGCTCTGAG	GACCAACAAG	AGCAGCCGGG	CCAGGGGCTG	24290
	GATGCAGCAC	GGCCGAGGTT	CCTGGATCCG	TGTCCTGCTG	TGGTGCCGAG	CCTCCGTGCG	CTTCCGCTTA	24360
	CGGGGCCCCG	GGACCAGGCC	ACGACTGCCA	GAGGCCACCC	GGGCTCTGAG	GATCCTGGAC	CTTGCCCCAC	24430
	GGCTCCTGCA	CCCCACCCCT	GTGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
50	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAAACAG	24570
	GGTCTGAGGA	AGCTGGGAGG	GGTCTAGGTT	CCCGGCTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
	TCTCCCTTGG	GTCCCTATGG	TGGGGTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
	CCCCGGCCAG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTTACGCT	GCTCAACTAC	GAGCGGGGCG	24780
	GGCGCCCGCG	CCTCCTGGGC	GCCTCTGTGC	TGGGCTGGGA	CGATATCCAC	AGGGCCTTGG	GCACCTTGTG	24850
	GCTGCGTGTG	CGGGCCAGG	ACCCGCCGCC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGCCGG	GGACCCCGCT	24920
55	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
	GGGTGGGCGG	CAGGGAGTGC	AGGTGACCTT	GTCACTGTTG	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTTACGCTTT	CCTGCAGCAC	ATGGGGCCGA	CTGTGCAACC	TGACTGCCCC	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTACAGAG	AGGAACCGCA	ACGGCTCAGC	CACCAAGGCC	CGGTGCCCTT	25200
	CACCCAGTCT	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
60	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTT	25340
	ATGTCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCCTCGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCG	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTTCAT	GCCAGCATCA	TCAAACCCCA	GAACACGTAT	25480
	TGCGTGGGTC	GGTATGCCGT	GGTCCAGAAG	GCCGCCCATG	GGCACGTCCG	CAAGGCCCTT	AAGAGCCACG	25550
	TAAGGTTTAC	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCAGTC	25620
65	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGTGGCA	25690
	CGTGTGTGTC	TGTTGTGCATG	TATCTGTGGC	TGTCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCTGTG	GTGTGCATGT	GTGTGTGTCT	GTGACACGTG	CATGTTTCATG	CTGTGTGCTG	25830
	CATGTCTGTG	ATGTGCCATAT	TTGTGGTGTG	TGTGTGCATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCTTGGC	CTTACTCCTT	CCTCCTCCAG	GCATGGTCCG	CACCATTTGC	CTCACGCTCT	25970
	CGGGTGCTGG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TTCTAGCATG	GGTGCCCTG	TCCGTGTCAA	26040
70	GGGCTGGGCC	TTGGAGACTG	TAAGCCAGGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGCGAC	26110
	CCCTGGCACC	CCCAGGACCC	CAGTCTGGCC	TATGCCCGGT	CCATGAGATA	TAGGAAGGCT	GATTTCAGGC	26180
	TGCGTCCCCG	GGACACACTC	CTCCAGAGC	GGCCGGGGGC	CTTGGGGCTC	GGCAGGGGTG	AAAGGGGGCC	26250
	TGGGCTTGGG	TTCCACCCCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGGCC	26320
	GGGTGCAGAG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCACAAGGA	26390
75	CCTCTTTCTC	TGACTTCTTG	AGCT					26414

Contig 2:

	197TGTGGGATTG	198GTTTTTCATGT	199GTGGGATAGG	200TGGGGATCTG	201TGGGATTGGT	202TTTTATGAGT	203GGGGTAACAC	20470
	197AGAGTTCAAG	198GCGAGCTTTC	199TTCCTGTAGT	200GGGTCTGCAG	201GTGCTCCAAC	202AGCTTTATTG	203AGGAGACCAT	204140
5	197ATCTTCCTTT	198GAACTATGGT	199CGGGTTTATA	200GTAAGTCAGG	201GGTGTGGAGG	202CCTCCCCCTGG	203GCTCCCTGTT	204210
	197CTGTATCTTC	198CACTCTGGGG	199TCGTGTGGTG	200CCTGCTGTGG	201TGTGTGGCCG	202GTGGGCAGGG	203CTTCCAGGCC	204280
	197TCCTTGTGTT	198CATTGGCCTG	199GATGTGGCCC	200TGGCTACGCT	201CCGTCCTTGG	202AATTCCCTCG	203CGAGTTGGAG	204350
	197GCTTTCTTTC	198TTTTCTTTTT	199TCTTCTTTT	200TTTTTTTTTT	201TGATAACAGA	202GTCTCGCTCT	203TTTTTGCCCA	204420
	197GGCTGGAGTG	198GTTTGGCGTG	199ATCTTGCGTC	200ACTGCAACCT	201GTGCTTCCTG	202AGTTCAAGCA	203ATTCTCTTGC	204490
10	197CTCAGCCTCC	198CAAGTAGCTG	199GAATTATAGG	200CGCCCACCAC	201CATGCTGACT	202AATTTTTGTA	203ATTTTAGTAG	204560
	197AGACGAGGTT	198TCTCCATGTT	199GGCCAGGCTG	200GTCTCGAACT	201CCTGACCTCA	202GGTGATCCTC	203CCACCTCGGC	204630
	197CTCCCAAAGT	198GCTGGGATGA	199CAGGTGTGAA	200CCGCCGCGCC	201CGGCCGAGAC	202TCGCTTCCTG	203CAGCTTCCGT	204700
	197GAGATCTGCA	198GCGATAGCTG	199CCTGCAGCCT	200TGGTGTCTGAC	201AACCCTCCGT	202TTCCTTCTCC	203AGGTCTCGCT	204770
	197AGGGGTCTTT	198CCATTTTCATG	199ACTCTCTTCA	200CAGAAGAGTT	201TCACGTGTGC	202TGATTTCCCG	203GCTGTTTCTT	204840
15	197GCGTAATTGG	198TGCTGTGCTG	199TTATCGATGG	200CCTCCTTCCA	201TTTCCTTTAG	202GCTTTGTTTA	203TTGTTGTTTT	204910
	197TCCGGCTCCT	198TGAAGGAAAA	199GTTTCGATTA	200TGGATGTGTTG	201AACTTCTTT	202TCTAAACAAG	203CATCTGAAGT	204980
	197TGCCGTTTTT	198CCTCTAAAGC	199AGGGATCCCG	200AGGCCCTGG	201CTGTGGAGTG	202GCACCGGTCT	203GGGGCTGTT	2041050
	197AGGAACCCGG	198CGCACAGCGG	199GAGGCTAGGT	200GGGGTGTGGG	201GAGCCAGCGT	202TCCCGCCTGA	203GCCCCGCCCC	2041120
20	197TCTCAGATCA	198GCAGTGGCAT	199CGGTGCTCA	200GAGGCGCACA	201CACCTACTG	202AGAAGTGTG	203GTGAGAGGGG	2041190
	197TCTAGATTCT	198GTGCTCCTTA	199TGGAATCTA	200ATGCCCTGATG	201ATCTGAGGTG	202GAACCGTTTG	203CTCCCAAAAC	2041260
	197CATCCCTTTC	198CCCCTGCTG	199TCTGTGGAA	200AAATCGTCTT	201CCACGAAACC	202AGTCCCTGGT	203ACCACAATGG	2041330
	197TTGGGGACCC	198TGTGCTAAAG	199ACCTGCTTCA	200GCAGCCTCTC	201GTCAGTGTG	202ATATATTGGC	203TTTTCTGTGT	2041400
	197TGAGTCCAGA	198ATAATTACGG	199ATTTCTGTGA	200TGCTTTCCCG	201CGACCTCAGA	202CCCATGGGCT	203ATTTGTGGGC	2041470
25	197GTGTTGCGCTG	198CTCCTGGGTT	199GGGAAGGGTG	200CAGGCCCCAT	201GTACCTTCCT	202GTTACTGCCT	203TCCAGGTTGG	2041540
	197TTCTCAGGGT	198TGAATCGTAC	199TEGATGTGGT	200TTTAGCCAC	201GGCCCTGCGG	202CCAGCTCCTG	203GGGGCTGGGG	2041610
	197AACATGCTGA	198AGCACAGAGT	199CACCGTCCGC	200GTCTTTTGAT	201GCCTCACAAAG	202CTCGAGGCCCT	203CTGTGTCCG	2041680
	197TGTTAGTGTG	198TGTCACGTGC	199CTGCTCACAT	200CCTGTCTTGG	201GGACGCAGGG	202GCTTAGCAGG	203TCCCGTAGTA	2041750
	197AATGACAAGC	198GTCTGGGGG	199AGTCTGCAGA	200ATAGGAGGTG	201GGGGTGCCGG	202TCTCTCTCCC	203GCGTCTTCAG	2041820
30	197ACTCTTCTCC	198TGCCGTGTCT	199GTGGCTGCAC	200CTGCATCCCT	201GCAATCCCTC	202CAGCACTGGG	203CTGAGAGGGC	2041890
	197CCGGGAGCTC	198GAGTGCCACT	199TGTGCCACGT	200GACTGTGGAT	201GGCAGTCCGT	202CACGGGGGTC	203TGATGTGTGG	2041960
	197TGACTGTGGA	198TGGCGGTTGG	199TCACAGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042030
	197TGTGTGACTG	198GTGGATGGCG	199GTCGTGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042100
	197TGTGTGACTG	198GTGGATGGCG	199GTCGTGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042170
35	197TGGTGAAGTG	198GGATGGCAGT	199CGTGGGGTCT	200GATGTGTGGT	201GACTGTGGAT	202GGCGGTCTGTG	203GGGTCTGATG	2042240
	197TGGTGAAGTG	198GGATGGCAGT	199CGTGGGGTCT	200GATGTGTGGT	201GACTGTGGAT	202GGCGGTCTGTG	203GGGTCTGATG	2042310
	197TGTGTGACTG	198GTGGATGGCG	199GTCGTGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042380
	197TGTGTGACTG	198GTGGATGGCG	199GTCGTGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042450
	197TGTGTGACTG	198GTGGATGGCG	199GTCGTGGGGT	200CTGATGTGTG	201GTGACTGTGG	202ATGGCGGTG	203TGGGGTCTGA	2042520
40	197GGTCTGACTG	198TGGTGAAGTG	199TGGATGGTGA	200TCGGTCTGATG	201GGGTCTGATG	202TGTGGTGAAG	203TGGGGTCTGA	2042590
	197GTCGTGGGGT	198CTGATGTGTG	199GTGACTGTGG	200ATGGCGGTG	201GTCCCGGGGG	202TCTGATGTGT	203GGTGACTGTG	2042660
	197GATGGCGATC	198GGTCACAGGG	199GTCTGATGTG	200TGGTGAAGTG	201GGATGGCGGT	202CGTGGGGTCT	203GATGTGTGGT	2042730
	197GACTGTGGAT	198GGCGGTCTGTG	199GGGTCTGATG	200TGGTGAAGTG	201GGATGGCGGT	202CGTGGGGTCT	203GATGTGTGGT	2042800
	197GACTGTGGAT	198GGCGGTCTGTG	199GGGTCTGATG	200TGGTGAAGTG	201GGATGGCGGT	202CGTGGGGTCT	203GATGTGTGGT	2042870
45	197GACTGTGGAT	198GGCGGTCTGTG	199GGGTCTGATG	200TGGTGAAGTG	201GGATGGCGGT	202CGTGGGGTCT	203GATGTGTGGT	2042940
	197GTGGTGACTG	198TGGATGGCG	199TGGTGAAGTG	200TGGTGAAGTG	201TGGTGAAGTG	202TGGTGAAGTG	203TGGTGAAGTG	2043010
	197GTGGTGACTG	198TGGATGGCG	199TGGTGAAGTG	200TGGTGAAGTG	201TGGTGAAGTG	202TGGTGAAGTG	203TGGTGAAGTG	2043080
	197ATGTGTGGTG	198ACTGTGGATG	199GCGGTCTGTG	200GGTCTGATGT	201GGTGAAGTG	202GGTGAAGTG	203GGTGAAGTG	2043150
	197ATGTGTGGTG	198ACTGTGGATG	199GCGGTCTGTG	200GGTCTGATGT	201GGTGAAGTG	202GGTGAAGTG	203GGTGAAGTG	2043220
50	197GGGTCTGATG	198TGTGGTGAAG	199GTGGATGGCG	200GTCGTAGGGT	201CTGATGTGGT	202GTGACTGTGG	203ATGGCAGTCG	2043360
	197GTCACAGGGG	198TCTGATGTGT	199GGTGAAGTG	200GATGGCGGTC	201GTGGGGTCTG	202ATGTGTGGTG	203ACTGTGGATG	2043430
	197GCGGTCTGTG	198GGTCTGATGT	199GTGGTGAAGTG	200TGGTGAAGTG	201TGGTGAAGTG	202TGGTGAAGTG	203TGGTGAAGTG	2043500
	197TGGCGGTCTG	198GGGGTCTGAT	199GTGGTGAAGTG	200TGGTGAAGTG	201TGGTGAAGTG	202TGGTGAAGTG	203TGGTGAAGTG	2043570
	197GCAGGTGGAG	198TCCCAGGTGT	199GTCTGTAGCT	200ACTTTGCCGT	201CTCGGCCCTC	202CGGCCCTCCT	203TTCCCAAAAC	2043640
55	197GAAGCTTCCC	198AGGCGTCTC	199TGGGCTTCAT	200CCCGCCATCG	201GGCTTGGCCG	202CAGGTCCACA	203CGTCTGATC	2043710
	197GGAAGAAACA	198AGTGCCAGC	199TCTGGCCGGG	200GCAGGCCACA	201TTTGTGGCTC	202ATGCCCTCTC	203CTCTGCCGGC	2043780
	197AGGTCTCTAC	198CTTGACAGAC	199CTCCAGCCGT	200ACATGCCACA	201GTTCTGGGCT	202CACCTGCAGG	203AGACCAGCCC	2043850
	197GCTGAGGGAT	198GCCGTCGTCA	199TCGAGCAGGT	200CTGGGCACTG	201CCCTGCAGGG	202TTGGGCACGG	203ACTCCAGCA	2043920
60	197GTGGGTCTCT	198CCCTGGGCAA	199TCACTGGGCT	200CATGACCAGA	201CAGACTGTGT	202GCCCTGGGGG	203GCAGTGGGGG	2043990
	197GAATGAGCTG	198TGATGGGGGC	199ATGATGAGCT	200GTGTGCCTTG	201GCGAAATCTG	202AGTGGGGCCA	203TGCCAGGCTG	2044060
	197CGACAGCTGC	198TGCAATTCAGG	199CACCTGTCTA	200CGTTTGACTG	201CGCGGCTCTC	202CTCCAGTTCC	203GCAGTGCCTT	2044130
	197TGTTCATGAT	198TTGCTAAATG	199TCTTCTCTG	200CAGTTTGATG	201CTTGAGGCCA	202AAGGAAAGGT	203GTCCCTCTCC	2044200
	197TTTAGGAGGG	198CAGGCCATGT	199TTGAGCCGTG	200TCCTGCCACG	201CTGGCCCTC	202AGTGTGGGT	203CTGAGGCCAA	2044270
65	197AGGAACAGTG	198TCCCTCTTCT	199TAGGAGGACG	200GGCCGTGTTT	201GAGCCACGCC	202CCGCTGAGCG	203GGCCTCTCAG	2044340
	197TGCTGGGTCT	198TCCACGTGG	199CCTGTGGCC	200CTTTGAGAGT	201GTGGTCTGTC	202CACGTGGGCC	203TGTTGGCTCTT	2044410
	197TGCAGATGCC	198TGTTAGCACT	199TGCTCGGCTC	200TAGGGGACAG	201TCGTGTCCAC	202CGCATGAGGC	203TCAGAGACCT	2044480
	197CTGGGCGAAT	198TTCTTGGCT	199CCAGGGTGG	200GGGTGGAGGT	201GGCCTGGGCT	202GCTGGGACCC	203AGACCTGTG	2044550
	197CCCCGAGCT	198GGGCAGCAAC	199TCCTGGATCA	200CATATGCCAT	201CCGGGCCACG	202GTGGCTGTG	203TGGGTGTGAG	2044620
70	197CCCAGCTGGA	198CCCACAGGTG	199GCCACAGAGA	200GACGTTCTGT	201GTCACACACT	202GTGGCTAAGC	203CCATGTGTGT	2044690
	197CTGCAGAGAC	198TCGGCCCGGC	199CAGCCACAGA	200TGCCCTGCA	201TTCCAGCCCA	202GCCCGCACT	203TCATCACAAA	2044760
	197CACTGACCCC	198AAAAGGGACG	199GAGGGTCTTG	200GCCACGTGGT	201CCTGCCTGTC	202TCAGCACCCA	203CCGGCTCACT	2044830
	197CCCAGTGTCT	198TCCCGTCTGC	199TTTCGACAG	200CTCCTCCCTG	201AATGAGGCCA	202GCAGTGGGCT	203CTTCGACGTC	2044900
	197TTCCTACGCT	198TCATGTGCCA	199CCACGCCGTG	200CGCATCAGGG	201GCAAGTGAGT	202CAGGTGGCCA	203GGTGCCATTG	2044970
	197CCCTGCGGGT	198GGCTGGGCGG	199GCTGGCAGGG	200CTTCTGCTCA	201CCTCTCTCT	202GCCCTTCCC	203CACTGNCCTT	2045040

	CTGCCCGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGCGGCA	GCCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
	AAGCGGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGCC	5250
5	GAGCGTTTGA	GCCTGCAGCT	TGTGAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTT	5320
	GTATCTCTCT	CTCCCGATAC	AAAAGGATTT	TATCCGATTG	TCATTCTCTG	CCCTGTCTGT	TGACCCCGCG	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTCCCATC	TGGAAGGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCGG	GGGGCTGTGT	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTACAC	CAGCCACTGG	5530
	GTGAGCCACA	CTCAGCGTGG	TAGAGCCACA	GTGCCTGGTG	CCACATCAGC	TCCTCTGGAT	TTTAAGTAAA	5600
10	ACCACACACC	TCCCAGGAGG	CATCTGCCTG	CGACCCCTGT	TGTGCCTGGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTCGTGCA	CACTCAAGGT	CATCAGCAAG	GTCCATCCGA	GTGAGGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
15	TGTGCACATG	TGTGTAAGCG	GCCCCCAGGC	CCACAGAAAT	CGCTGACAAA	GTCACCTCCC	CAGAGAAGCC	6020
	ACCACGGGGG	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAAATGTGAGG	TGATGACTGC	GTCTCATGCG	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	CTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAAATC	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GCCGAGTCCC	TAGAGTTGGT	CGGTGTGCTT	CTGTGCAAAA	6370
	AGTGCAGTCC	TCTTGCCCAT	CACTGTGATA	TCTGCACCAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTTCTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACTGTTG	TCTGCCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCTGCTCG	CAGCCTCCCG	AGCAGCTGAG	ATTACAGGCA	6580
25	CCCACCCCTC	GCGCCTGGCT	AATTTTTGTA	TTTTTAGTAG	AGAGGGGTTT	TTGCCATGTT	GGCCAGGCTG	6650
	GTCTCGAATT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCACGCC	CAGCCGGAAA	GCCTCTTTTT	AAGGTGACCA	CCTATAGCGC	TTCCCGAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	TAGAGGTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTT	GTGGGCTCTG	GGGATGCTGT	6860
	AGGGTCCGGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCCACGGG	GCTATTCTGC	TCTCACTGTT	6930
30	TGTCTGAAAA	CGACCCCTTG	GCATCCTTTG	TTGGAGAGTT	TCTGCTTCTC	GTTGGTCATG	CTGAAACTAG	7000
	GGGCAAGGTT	TATCCGTTTG	GCGCGCAGCG	GCTACATGTA	GGGTCAATGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAAGGA	GTCCGGTTAA	GCATTCTATC	CGGGTCAAGT	GTCTGGTTCT	GTGAATAAAC	7140
	TCTAAGATTG	AAGAAACCTT	AATGAAAGAA	AACCTTGATG	ATTACAGACA	AGGATGTGGT	CACACCTGTG	7210
	GCTGCAATCTG	TTTACGCCGC	CCCAGTGCA	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGTG	CAGAGGTCTC	7280
	ATCTGGTATG	TTTCTGAGGT	GTTTGCCGGC	TGAATGGTAG	ACGTGTGCTT	TGTGTGTATG	AGGTTCTGTG	7350
35	TCTGTGTGTG	GCTCGGTTTG	AGTGTACCGA	GTGCCAGCAG	ATGCCCTGCC	CGTCTCTCAC	CTGTGTCTTC	7420
	CCGCCCCAGG	TCTTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTCAGCT	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGCGGGGA	TTGCGCGGGA	CGGGTGAGGC	CTCCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTTGATTTGC	TTTTGATGCA	TTTCACTGTA	ATATTCCTGG	TGCTCTGGAG	7630
40	ACCATGACTG	CTCTGTCTTG	AGGAACCGA	CAAGGTTGCA	GCCCCCTCTT	GGTATGAAGC	CGCAGGGGAG	7700
	GGGTTGCACA	GCCCTGAGGAC	TGCGGCGCTC	ACGCAGGCTC	TGTCCAGCGG	CCATGTCCAG	AGGCCCTCAGG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCACACCAG	7840
	CTTCTGTCTC	GTCACCCAGG	TTCCGTTAGG	GTCTTGGGG	AGATGGGGCT	GGTGACGCTT	GAGGCCCCAC	7910
	ATCTCCAGCT	AGGCCCTCGA	CAGGTGGGCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCA	TCCCACTTGC	7980
	ATGGGGTCTA	CACCCAAAGG	CGCACACACC	TAAATATCTG	GCCAACTTAA	TGTGGTTCAA	CTCAGCTGGC	8050
45	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGACG	8120
	GTAGTTTACA	TAGTATATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCTCTA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCCCACC	8260
	CTGTGTCCAA	GTTTCTCTAT	TGTTCAAGTT	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTTTCTCTTC	8330
	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCTG	CAATGTCCCT	ACAAAGGACA	TGAACCTATC	8400
50	CTTTTTTATG	ACTGCATAGT	ATTCCGTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTGCG	GTTGGTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCGCAAT	AAACATACGT	GTGCAATGTG	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAAAATGGAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCAGCA	CACCTGTCTC	CACAAATGTT	GAACATAGTT	ACACCTCCAC	8680
	CAACAGTGTA	AAAGTGTCTT	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
55	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTCTGT	CATCTTTTGA	AACCTTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCTAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTTGATCA	8960
	CACAACCTGC	TCTGGGATTT	GGAGGAAAGT	GTCTCTGAGC	TGGCGGCACA	CTGGTACGCC	CTCTGGGACA	9030
	GGATACTCTT	GGCCCATGGT	CATGGGGCCG	TGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCACG	9100
60	CTTCTGTGCG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTTCCACAG	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCCTAAGAG	9240
	TCTGAGAAGT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
	AAAGAAATTT	AAGTTTTCAT	TTTAAACCGT	TTGGAGAAATG	TTACTTTATT	TATGGCTGTG	TAAATGTTT	9450
65	GACATTCAGT	CCCTCGTAGA	CAGATACATC	GTAAGAAAGT	TAAAGTTAAT	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCTGCGT	TTGGTGGATG	ATTCTTGTGT	GGTGACACCT	CACCTCACCC	ACGCCAAAAAC	9590
	CTTCTCTCAG	TGAGGCCCGT	GCCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTGTGC	ACCGCAGCGT	TGTCTCTGCC	AAGTCTCTCT	TCTCTGCCGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCCGTTGG	CCGTGCACCC	AGGCCTGGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
70	ACCGTGACAG	CCCTGGTCTT	GCAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGTGCT	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCTCT	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCTG	CAGGGCCGAG	GCGGCAGCTG	CCTCCCCAGG	GTGCACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGGCCACAC	GCCTTCGCTG	CGGTACCGTT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGTG	CTGCTGTCTT	GAACACAGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTGAATC	AAACTAAAAT	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTCCA	ATGTGGTCCC	10220
	CCTCAAGGGC	GCCCCACAGA	GCCGCTGGGC	TTGTTTAAAA	GTGCGATTGG	ACGAGGGACG	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAAATGT	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTGTGAAAAC	CCATTTGGAC	CCGCCCTCCA	AGTCCACCCT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGG	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCGGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTCCT	GAGGGTGCTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGGAA	TGCACCAGGG	GAGCTGCGCA	GCTGGCCGAG	GTCCACAGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCGCA	GGGAAGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCGA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAACT	10990
10	CAGAACCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCCTT	AGGGCATCTA	GGAGAAAAAC	GGCAAAAGTCG	11060
	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTTAGTCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGTT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCCTGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGCATGC	TCACCTACCT	11270
	GTCCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGCCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTTCTCT	GGGGTTTTTC	AAAGCATTTA	ACAAGGGTGT	11410
15	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCTGGTCT	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCGTGGTG	AACTTGCAGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCATTTC	CCCTGGTGCG	GCCTGTGTGT	11620
	GGATACCCGG	ACCCTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCAAC	TGGCCGGAAG	TGGAGCCTGT	11690
20	GCCCCGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTGCGTCC	ACCTCTGCTT	CCGTGTGGGG	CAGGCCAGTG	11760
	CCAAATCCCAA	AGGGTCAGAG	GCCACAGGGT	GCCCTCTGTC	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAGAGGAA	ATGGTGCACC	11900
	AGACTTGGGT	GCACTGAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCTGTGCTG	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GCCCTGTCTG	GCATGAGTCC	CTCTGAACCC	GAGACCCTGG	12040
25	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
	GTGAGCCCCA	CACCTCAAGG	CTCATCCACA	GCTCTACAGG	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTCCTC	CAGAGCCCCG	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAAATGAAT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCAGGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCCGAG	12390
30	ACTTTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAACC	AACATAGTGA	12460
	AATTCATTTT	CTACTTAAAA	AATACAAAAA	TTAGCCTGGC	CTGGTGGCAC	ACGCCCTGTG	TCCCGGCTAT	12530
	GCGGGAGGCT	GAGGCAGGAG	AATCATTTGA	ACCCAGGAGG	CAGAGTTGTC	AGTGAGCCGA	GATCACACCA	12600
	CTGCACTCCA	GCCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAAGTAT	AGCAATCCAA	TAGAGGCCCG	12670
	AACCATAGTG	GACAGGTGTT	TTTTTATTCT	GTCTTTCGAT	ATATTTTACT	GGTGTGTGTC	TAGAGGCCCG	12740
35	AACTGGGGGT	GCCTTCTCTT	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTAAAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTCTGTT	TGAGTTAACA	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGTCT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGACGCCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGGCCAGC	ATGTCCCTGT	13020
	TGAGGCCCTC	TCCCAACAAG	GATGCCGGTC	TCTGTGTGTC	CCCACAGTCC	CTGCTTCCCT	CTCACAGGCT	13090
40	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGCATGATTT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCTCTG	CAGTGTCTGG	CATACCAAGT	AGCTGTGAAC	TGTCCACTGC	TTATTTTGCT	13230
	CCCATGAAA	TGATTTTTTT	AGGACAGGCA	CCCCGTGTTT	CAGCCTCTGG	CACAGCATCA	TGAATGTGTA	13300
	TTGAAGGACA	AAGGACAGAC	AAACAAATCA	GGAAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGGTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAAATATC	TGTGTCCCA	13440
45	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAAGCTCA	GCAGTGAGG	CATGTGGTTC	13510
	CCATACTCAG	GGTGAATCTA	CATCCTCTGT	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CATCTGGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAGAAAAGT	GAAAAAGGAA	AGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAACAC	AGCTCAGATG	GTAAGATGTG	GTGAGAACTG	ATGGACAGAA	CAATAGAACA	13720
	AAACGGAAAG	CTATCTCTC	AGAAACGTGT	GTTAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTTTCT	GAGAAAACCTG	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACAAC	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	GTGTGGGGA	AGGACACACA	GGGAGGCGGA	TGAACCAAGT	14000
	GAGGCAACGG	GCATTGCTTT	CACTGCAGAG	AAACTCAGCT	TGCCTGAGCC	ACAGTGAAAA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGGCCCTTGC	GAGGTCTGTC	ACATTCTATC	TCTCACTTTG	14140
55	TTCTCTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGTT	CACCCAGTGT	14210
	GCAAAGGGCA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCCCTTGC	TCTGCCCAG	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTGGG	GATGGCTGTG	AAAGAAAGAA	14350
	TGGACGTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTTGAGGCAA	GCTGGAAGAA	CTTTTCTGGA	AAGCAGCTTG	14560
	TTTGCAATGA	AGTCCTCACA	ATGTCCTGTG	TCTTCCCACT	AATTCACCTT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTTACCAAT	TCCAGTGTTT	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTG	CTCTGCCATT	AAACATTTTT	CAAGAATTTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTT	AAAGCTGGAT	GTAAAGAAAC	14840
65	ACACCCAGGA	AGCCTGCCGT	GAATGTCATG	TGTGTTCTAT	TTTGGACATG	GACATACATG	GGAAGTGAAT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCCAT	CCTGCCCTCT	TGGAGACACC	ATGTGTGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTCTC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCGAGGGAC	15120
	GCCCGGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
70	GATGATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTGGGATA	CACTCAACAT	CACTAGCCAG	GTCTGGTGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCAATTTG	AGTCCATGGA	GTGAGCACC	AGCCCTCTCG	15330
	GGGTGAGCGC	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GGCTCTTTGG	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTG	GGGCAGGCAC	CTGTGTCTGA	CATTCCCCCC	TGTGTCTCAG	CTATGCCCGG	15470
	ATCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
75	TTGCGGTCTT	CGCGGTGAAG	TGTCAACAGC	TTTCTCTGGA	TTTGCAGGTG	AGCAGGCTGA	TGGTGCACAT	15610
	AGAGTTCAGA	GTTCAAGGAG	TGTGTGCGCA	AGTATGTGTG	TGTGTGTGTG	CGCGCGTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCACA	TGTACGCATA	TACACGTGAG	CACATACATG	TGTGCATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCACAAGTG	TGTGCACATG	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTTC	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCTCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCAGGGC	TTGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
5	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCCTGTGG	GCTTCTGTGT	CCACTCCCC	16100
	TCTCCTGTGG	GCATTTACAT	CCACTCCACT	CCCTCTCTCG	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	GTGGCATCTG	GGGTCCACCT	CCCCTCTCTG	TGGCAGTTTG	CGTCCACTCC	CTCTCCTGGT	TCCTTCTGT	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTGCCG	16310
	GGTGAGGGCC	AGGCCCGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCTCT	TTTCTTGTTC	16380
10	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGGCG	ACTCTAGGTG	AACAGCCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGTCTGA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCACGC	CTCCAGGGAC	CCTCCGCGCT	CTGCTCACCT	16590
	CTGACCCGGG	GCTTACACCT	GGAACCTCTG	GGTTTTAGGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCCGTG	GCACAGTTCT	GTTGCGGTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
15	AGGAGCCGGT	GTGGCCCGAG	GTGTCCCCAC	TGTGCCTGTG	CACTGGCCGT	GGGACGTCAT	GGAGGCCATC	16800
	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGTGT	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTTGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCCACACA	GCCCGGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCCT	GGAACGTTCC	CTGTCTGTGG	TGGTCAGGGG	GTGCCCTCTG	17080
20	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGCCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCTGG	17150
	AGTCAGGGCA	GGTGGTGCCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGCACCACA	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCCTCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCATAGAGG	17290
	GCTGAGAAGG	AGTGTAGACA	TTTGTGTTAC	CCAGGGCCGA	GGCTGCGCGA	ATTACCGTGC	ACACTTGATG	17360
	TGAAATGAGG	TCGTGCTGTA	TCGTGGAAAC	CACAGAAAGG	CTCACGGGAG	AGTTTTCAT	TACAAGGTCT	17430
25	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTGCGCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
	TGCATGTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCCAG	GSTCGCTCCG	17570
	GCTCAGACCG	CCCTCTCTCT	TGCCCTTCTC	CTCTGCTCCA	AATCTTCCCT	CGTTTGCATC	TCTCGTACGC	17640
	GTGCCTGGGC	CCTCGTGCAA	GCTGCTTGAC	TCCTTTCCGG	AAACCCTTGG	GGTGTGCTGG	ATACAGGTGC	17710
	CACTGAGGAC	TGGAGGTGTC	TGACACTGTG	GTTGACCCCA	GGGTCCAGCT	GGCGTGCTTG	GGGCCTCCTT	17780
30	GGCGCATGAT	GAGGTACAGG	GAGTTTTCCT	AGGTGAAAC	TCCTGGGAAA	CTCCAGGGC	CATGTGACCT	17850
	GCCACCTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGCTCTCA	TTTCCCCACC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCTGAT	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	GTGCTTGATG	TGGGTAGCCC	TGAGGAGGCC	GAGATGCCAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCGGA	18130
35	TTTCACGGCA	GCCAGGCTGC	AGTGGGCGAG	GCTGTGGTGG	TCCACGTGGC	GCTGGGGGGC	GGGTCTGATT	18200
	CAAACTCCGT	GGGGCTCGGC	CTTCTTGCCC	CGTGTGGGCA	GCGCCTCCAC	ACGGGCTTGG	GGTGGAGGCT	18270
	CCGACCTCTA	GCAAGTGGCT	ATTCTCCCT	TTGGAAGAGA	GCCCCCACC	CATGCTAGGT	GTTCCTCTCC	18340
	TGGGTACAGG	GCGTGGCCGT	GTGGCAACCC	CGGACCTTAA	GGCTTATTTA	TTTGTTTAAA	AACATTCTGG	18410
	GCCTGGCTTC	CGTTGTTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGTGCTG	GCTTGACTGG	TGTGATCTGA	GGTCATTCCA	GAAGTGGCTC	AGGAAGTCAG	TGAGACAGG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGC	TCAGGCACCTG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAGGTACA	18690
	CGGGGGCTCT	GACATACAGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGTGCTGT	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGGTGGCT	18830
45	CACACCTGTA	GTCCACAGCA	TTTGGGAGGC	CGAGGCGAGA	GGATCCCTTG	AGCCCAGGAG	TTTAAGAGTA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCATCTCT	ATGAAAAATA	AAAACAAAAA	TTAGCTGAAC	ATGTTGGTGT	18970
	GCGCCTGTAG	TTCCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAAGCTGCA	19040
	GTGAGCTGAG	ATTGCACCA	TGTACTGCAG	CCTGGGTGAC	AGAGTGAAG	CCCATCTCAA	CAACAACAAA	19110
	GAAGACTGAC	AAATGCAATT	TCTTGGAAG	AAACATTTC	TAGGAACCTA	ACCTACACAC	AGAAGCCAAG	19180
50	TCGGTGTCTC	GGTGTCAAGT	AGATGAGATG	TGCGGCTCTC	ACACCATCAC	CCAGACCCA	GGGTTATATG	19250
	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTTGTCTGAC	19320
	GAAGGGCAGG	ATTGATGATA	AGTACCTGCT	GGTACACAA	GAACAATGGA	TAAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTGACGCC	19460
	TCCACATGCG	TGTTTCATACA	GATGGTGCAC	AGAAACGCAG	TGTACCTGTG	CACACACAGA	CACGCAGCTA	19530
55	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGCATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCTG	TGGGCCCATG	CCACACCCCA	CAGCACCTGT	19670
	CTGATTAGGA	GGCCTTTCTC	CTGACGCTGT	CCGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTTCAT	AGCAAGTTTG	GAAAGACCCC	ACATTTTTC	TGCGCGTCAT	CTCTGACACG	GCCTCCCTCT	19810
	GCTATTCAT	CCTGAAAGCC	AAGAAGCCAG	GATGTGCAG	GTGCTGGCC	TCAGTGGCAG	CAGTGCCTGC	19880
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	TCTTACCCTT	TTTGCATCA	19950
	GGAAGTGGTT	TAACCCAACC	ACTGTGAGCC	TCGTCTGCCC	GCCCTCTCGT	GGGGTGGAGC	GAGCACCTGA	20020
	TGGAAGGGAG	AGGAGCTGTC	TGGGAGCTGC	CATCCTTCCC	ACCTTGTCTT	GCCTGGGGAA	GCGCTGGGGG	20090
	GCTTGGTCTC	TCTGTTTGG	CCCATGGTGG	GGTTTGGGGG	GCTTGGCCTC	TCTGTTTGG	CCTGTGGTGG	20160
	GATTGGGCTG	TCTCCCGTCC	ATGGCACTTA	GGGCCCTTGT	GCAAACCCAG	GCCAAGGGCT	TAGGAGGAGG	20230
65	CCAGGCCAGG	GCTACCCAC	CCCTCTCAGG	AGCAGAGGCC	GCGTATCACC	ACGACAGAGC	CCCGCGCCGT	20300
	CCTCTGCTTC	CAGTCAACCG	TCCTCTGCC	CTGGACACT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTCA	AGCCATGTCC	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTC	TTTCTGTGTT	20440
	GTGGAAATTT	CACCTGGAGA	AGCCGAAGAA	AAATTTTGTG	TCGTGACTCC	TGCGGTGCTT	GGGTGGGGAG	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGTGTGGG	CAGCTTTCCG	GTGTCTCTGT	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCCCCCA	GGGATGTGCG	TGGGGGCCAA	GGGCGCCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCAGTGG	CTGTGCCACC	AAGCATTCCT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTCACTTA	CGTGCCACTC	CTGGGGTCA	TCAGGACAGG	CAAGTGTGGG	TGGAGGCCAG	TGCGGGCCCC	20790
	ACCTGCCAGG	GGGTATCCTT	TGAACGCCCT	GTGTGGGGCT	AGCAGCCTCA	GATGCTGCTG	AAGTGCAGAG	20860
	GCCCCCGGGC	CTGACCCCTG	GGGCCCTGGG	CCACGCTGGC	AGCCCTATGT	GATTAACAGC	TGGTGTCCCC	20930
	AGGCCACGGA	GCCTGGCAGG	GTCCCCAACT	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGG	GATGGCTCCC	21000
75	CACGCTTGGG	AGCCTTCTGA	CCCCTGACCT	GTGCTCTCTC	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21070
	TCCTGGGGTC	CTGAGCAAGT	TCTCTCCCCG	CCCCGCCGCT	CCAGCGTCAC	TGGGCTGCCT	GTCTGCTCGC	21140
	CCCGGTGGAG	GGGTGCTGCT	CCCTTCACTG	AGTTTCCAC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21210

	TGCCCCGCCA	CCCACACGTC	CTAGGAGGGT	TGGAGGATGC	CACCTCTGGC	CTCTTCTGGA	ACGGAGTCTG	21280
	ATTTTGGCCC	CGCAGCCCAG	ACGCAGCTGA	GTCGGGAGCT	CCCCGGGACG	ACGCTGACTG	CCCTGGAGGC	21350
	CGCAGCCAAC	CCGGCACTGC	CCTCAGACTT	CAAGACCATC	CTGGACTGAT	GGCCACCCGC	CCACAGCCAG	21420
5	GCCGAGAGCA	GACACCAGCA	GCCCTGTAC	GCCGGGCTCT	ACGTCCCAGG	GAGGGAGGGG	CGGCCACAC	21490
	CCAGGCCCGC	ACCGCTGGGA	GTCTGAGGCC	TGAGTGAGTG	TTTGGCCGAG	GCCTGCATGT	CCGGCTGAAG	21560
	GCTGAGTGTC	CGGCTGAGGC	CTGAGCGAGT	GTCAGGCCAA	GGGCTGAGTG	TCCAGCACAC	CTGCCGTCTT	21630
	CACCTCCCCA	CAGGCTGGCG	CTCGGCTCCA	CCCCAGGGCC	AGCTTTTCCT	CACCAGGAGC	CCGGCTTCCA	21700
	CTCCCCACAT	AGGAATAGTC	CATCCCCAGA	TTCGCCATTG	TTCACCCCTC	GCCCTGCCCT	CCTTTGCCTT	21770
10	CCACCCCCAC	CATCCAGGTG	GAGACCTGA	GAAGGACCCT	GGGAGCTCTG	GGAATTTGGA	GTGACCAAAG	21840
	GTGTGGCCCTG	TACACAGGCG	AGGACCCCTG	ACCTGGATGG	GGGTCCCTGT	GGGTCAAATT	GGGGGGAGGT	21910
	GCTGTGGGAG	TAAAATACTG	AATATATGAG	TTTTTCAGTT	TTGAAAAAAA	TCTCATGTTT	GAATCCTAAT	21980
	GTGCACTGCA	TAGACACCAC	TGTATGCAAT	TACAGAAGCC	TGTGAGTGAA	CGGGGTGGTG	GTCAGTCCGG	22050
	GCCCCATGGCC	TGGCTGTGCA	TTTACGGAAG	TCTATGAGTG	AATGGGGTTG	TGGTCAGTGC	GGGCCCATGG	22120
15	CCTGGGTGGG	CCTGGGAGGT	TTCTGATGCT	GTGAGGCAGG	AGGGGAAGGA	GGGTAGGGGA	TAGACAGTGG	22190
	GAGCCCCCAC	CCTGGAAGAC	ATAACAGTAA	GTCAGGCCCC	GAAGGGCAGC	AGGGATGCTG	GGGGCCAGC	22260
	TTGGGGCGCG	GGGATGATGG	AGGGCCTGGC	CAGGGTGGCA	GGGATGATGG	GGGCCCCAGC	TGGGGTGGCA	22330
	GGGGTGATGG	GGGGGGCTGG	TCTGGGTGGC	GGGGAAGATG	GGGAAGCCTG	GCTGGGCCCC	CTCCTCCCCCT	22400
	GCCTCCCACT	TGCAGCCGTG	GATCCGGATG	TGCTTCCCTG	GTGCACATCC	TCTGGGCCAT	CAGCTTTTCAT	22470
20	GGAGGTGGGG	GGCAGGGGCA	TGACACCATC	CTGTATAAAA	TCCAGGATTC	CTCCTCCTGA	ACGCCCAAC	22540
	TCAGGTTGAA	AGTCACATTC	CGCCTCTGGC	CATTCTCTTA	AGAGTAGACC	AGGATTCTGA	TCTCTGAAGT	22610
	GTGGGTAGGG	TGGGGCAGTG	GAGGGTGTGG	ACACAGGAGG	CTTCAGGGTG	GGGCTGGTGA	TGCTCTCTCA	22680
	TCCTCTTATC	ATCTCCAGT	CTCATCTCTC	ATCCTCTTAT	CATCTCCAG	TCTCATCTGT	CTTCTCTTAA	22750
	TCTCCAGTC	TCATCTGTCA	TCCTCTTACC	ATCTCCAGT	CTCATCTCTT	ATCCTCTTAT	CTCCTAGTCT	22820
25	CATCCAGACT	TACCTCCAG	GGCGGGTGCC	AGGCTCGCAG	TGGAGCTGGA	CATACGTCCT	TCCTCAGGCA	22890
	GAAGGAATG	GAAGGATTGC	AGAGAACAGG	AGGGGCGGCT	CAGAGGGACG	CAGTCTTGGG	GTGAAGAAAC	22960
	AGCCCCCTCT	CAGAAGTTGG	CTTGGGGCAC	ACGAACCCGA	GGGCCCTGCG	TGAGTGGCTC	CAGAGCCTTC	23030
	CAGCAGGTCC	CTGGTGGGGC	CTTATGTTAT	GGCCGGGTCC	TACTGAGTGC	ACCTTGGACA	GGGCTTCTGG	23100
	TTTGAGTGCA	GCCCGGACGT	GCCTGGTGTG	GGGGTGGGGG	CTTATGGCCA	CTGGATATGG	CGTCATTTAT	23170
30	TGCTGCTGCT	TCAGAGAAATG	TCTGAGTGAC	CGAGCCTAAT	GTGTATGGTG	GGCCCAAGTC	CACAGACTGT	23240
	GTCGTAAATG	CACCTCTGGTG	CCTGGAGCCC	CCGTATAGGA	GCTGTGAGGA	AGGAGGGGCT	CTTGGCAGCC	23310
	GGCCTGGGGG	CGCCTTTGCC	CTGCAAACTG	GAAGGGAGCG	GGCCCGGGCG	CCGTGGGGCG	ACGACCTCAA	23380
	GTGAGAGGTT	GGACAGAAACA	GGCGGGGAC	TTCCAGGAG	CAGAGGCCGC	TGCTCAGGCA	CACCTGGGTT	23450
	TGAATCACAG	ACCAACaGGT	CAGGCCATTG	TTCAGCTATC	CATCTTCTAC	AAAGCTCCAG	ATTCTGTGTT	23520
35	CTCCGGGTGT	TTTTTGTGTA	AATTTTACTC	AGGATTACTT	ATATTTTTTG	CTAAAGTATT	AGACCCCTAA	23590
	AAAAGGTATT	TGCTTTGATA	TGGCTTAACT	CACTAAGCAC	CTACTTTATT	TGCTGTGTTT	TATTTATTAT	23660
	TATTTATTAT	ATTAGAGATG	GTGTCTACTC	TGTCACCCAG	GTTGTTAGTG	CAGTGGCACA	GTCTGGCTC	23730
	GCTGTAGCCG	CAAAACCCCA	GGCTCAAGTG	ATCCTCCGGC	CTCAGCTTCC	CAGAGTGCTG	GGATTACAGG	23800
	TGTGAGCCAC	TGCCCTTGCC	TGGCACTTTT	AAAAACCACT	ATGTAAGGTC	AGGTCCAGTG	GCTTCCACAC	23870
40	CTGTCACTCC	AGTAGTTTGG	GAAGCCGAGG	CAGAAGGATT	GTCTGAGGCC	AGGAGTTTGA	GACCAAGCATG	23940
	GGTAACATAG	GGAGACCCCA	TCTCTACAAA	AAATGCAAAA	AGTTATCCGG	GCGTGGGGTC	CAGCATCTGT	24010
	AGTCCAGACT	GCTCGGGAGG	CTGAGTGCGA	GGATCGCTTG	AGCCCGGGAG	GTCATGGCTG	CAGTGAGCTG	24080
	TGATTGTACC	ATCGCACTCC	AGCCTGGGCA	ACAGAGTGAG	ACCCTGTCTC	AAAAAAAAAA	AAAAAAAAAAG	24150
	AAGGAGAAGG	AGAAGAGAAG	AAGAAGGAAG	AAGGAAAGAG	AAGAAGAAGG	AAGAAGGAAG	AAAGAAGGAG	24220
45	AAGGAGGCC	GCTAGGTGCT	AGGTAGACTG	TCAATCTCA	GAGCAAAATG	AAATAAACAA	AGTTTTAAAG	24290
	GGAAAGAAAA	ACCCCACTC	TTTGGACTTC	CTTAGGCCCTG	AACTTCATCT	CAAGCAGCTT	CCTTCCACAG	24360
	ACAAGCGTGT	ATGGAGCGAG	TGAGTTCAAA	GCAGAAAGGG	AGGAGAAGCA	GGCAAGGGTG	GAGGCTGTGG	24430
	GTGACACCAG	CCAGGACCCC	TGAAAGGGAG	TGGTGTGTTT	CCTGCCTCAG	CCCCACGCTC	CTGCCGGTCC	24500
	TGCACCTGCT	GTAACCGTCG	ATGTTGGTGC	CAGGTGCCCA	CCTGGGAAGG	ATGCTGTGCA	GGGGGCTTGC	24570
50	CAAACTTTGG	TGGGTTTCAG	AAGCCCCCAGG	CACCTGTGGC	AGGCACAATT	ACAGCCCTC	CCCAAGATG	24640
	CCCACGTCTCT	TCTCTGGAA	CCTGTGAATG	TGTCACCCGC	AAGGCAGAGG	CTGGTGAAGG	CTGCAGGTGG	24710
	AATCACGGCT	GCCAGTCAGC	CGATCTTAAG	GTCATCTCTG	ATTATCTGGT	GGGCCCTGATA	TGGCCACAAG	24780
	GGTCCTTAGA	AGTGAGAGAG	GGAGGCAGGG	GAGAGCTCAGA	GAGGGGACGT	GAGAAGGACC	ATTGGCCACT	24850
	GCTGGCTTTG	AGATGGAGGA	GGGGTCCCC	AGCCAAGGAA	TGGGGGCAGC	CGCTCCATGC	TGGAAAAGCA	24920
	AGCAATCCTC	CCCGGTCCTG	AGGGCACACG	GCCCTGCCCA	CGCCTCGATT	TCAGGCCAGT	GGGACCTGTT	24990
55	TCAGCTTTTC	GGCCTCCAGA	GCTGTAAGAT	GATGCGTTTG	TGTTCAAGCCA	CTAAGCTGCA	GTGATTGCTC	25060
	ACAGCAGCAA	ATGGAATAGC	AGTACAGGGA	AATGAATACA	GGGACAGTTC	TCAGAGTGAC	TCTCAGCCCA	25130
	CCCCTGCG							25138

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in

5 den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon			Intron				3'-Exon			
Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N C	A G G
10 Häufigk.(%)	70	60	80		100	100	95	70		80	100 100 60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616

15 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

10 Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position 21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGACACGGTGATCTCTGCC
TCTGCTCTCCCTCCTGTCCAGTTTGATATAAATTACGAGGTTACCTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGTTGCCGGCAATGGGGAGAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA
CGAGAACCCCTCTTCTGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA
AGATTTAATGTGTGTTGACGGCCAGGTGCGGTGGCTCAGCCGGTAATCCCAGCACTTTGGGAAGCTGAGGCAAGTGA
TCACCTGAGGTGAGGAGTTTGAGACCAGCCTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAAATTAGCTG
15 GGCATGGTGGTGTGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCTGGGCGACAAGAGTGAACTCTGTCTTTAAAAAAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCCTCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGTGGTGTTCAGGGG
ATGGTGCTGCTGGGCCCTGCCGTGTCCCCACCCTGTTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCCTT
20 TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTACCCCCCTCCCAAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTTCTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGAGAATTGCTAACTCGGCGGTGTTTA
CAGCAGGTTGCTTGAAATGCTGCGTCTTGCCTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCCGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTGTGCTCCAGCTTCTCTCGTTGAG
25 GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCTTGTGTTCTCTGACTTCAGATGAGGTGACAATCTGCCCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGCCCCGGGTGTCCCTGTCCGTGCAGCGTGATTGAGGTGTGGCCCCGGGTGT
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGGTGCCCCG
GTGTCCCTGTCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG
CGCGCCCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTTGCCCATTGCTGGGTAGATGGTGACGGCGCAGTGCTGGTCCCCAAGCCTATCTTTTCTGATGCTCGGCTCT
35 TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCCTGGCTTGCTCACCACGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCAGCTTGTCTCATGCCGAGGCTGGACTCTGGGCTGCCTGTGTCTGCTGCCACGTGTTGCTGGAGACATCCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGCTCTGTTTTCTCCCAAGCTGCCCTCTGC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGCTCACCTTATCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCCT
40 CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTGGAGTGTCTGTCTGTCTCTGCTCTGAGACCCACGTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTCAGACTTCCCTCTTGGGTCTTAGTTTTGAATTTCACTGATTTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTCTTGGTTTATTCTTTCATTCCCTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTTCTTAA
5 AACTTTCAAAGTGTTAACTTCTTTTAAGTATTCTTATTCTGTGATTTTTTCTTTGTGCACGCTGTGTTTTGACGTGA
AATCATTTTGATATCAGTGACTTTTAAGTATTCTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATAACCAATTATTTGAAGTTTGCGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTTCTAGATGCATGAAATTCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTTCAAT
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA
TCTTTTGAGACTTCTATGTCTCTAGTAATCTAGTAATTTCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTTCTGCTGTGTCTGTTTTCTGCCTTTAATTTATATATATATATATATATTTTTTTTTTTTGGAGACA
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGAGTGGTGTGATCACAGGTGAGTGAACTTTACCTTCTGGCCTGAGCCGT
CCTCTCACCTCAGCCTCTGAGTAGCTGGAACGCGACACGACCGCTACACCTGGCTAATTTTTAAATTTTTTCTGGA
15 GACAGGGTCTTGCTGTGTGCCCAGGCTGGTCTCAAACCTCTGGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTGTGA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCTGTGTTTAACTGGATAACCTGATTTATTTTCAATTTTTGTG
ACTAGAGACCCGCTGGTGCCTCTGATTCTCCACTTGCTGTGATGCTCCTCGTTCCCTTGTCTCACCACCTCTTG
GGTTGCCATGTGCGTTTCTGCGGAGTGTGTGTGATCCTCTCGTTGCTCCTGGTCACTGGGCATTGCTTTTATTCT
20 CTTTGCTTAGTGTTACCCCTGATCTTTTATGTGCTGTGTTGCTTTTGTATTAGAGACAGTCTCACTCTGTCAACCA
GGCTGGAGTGTAATGGCACAATCTCGGCTCACTGCAACCTCTGCTCCTCGGTTCAAGCAGTTCTCATTCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCACCACGCTGGCTAATTTTTGTATTTTGTAGTAGAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACCTGACCTCAAGTGATCTGCCGCTTGCTGCTCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCTTGTCTGAGCAATAAGACCTT
25 AGTGATTTTAGCTCTGGCCACCCCTGAGCTGTGTGCTGTTTTCCCTGCTGACTTAGTCTATCTCAGGCATCTTGACA
CCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTGAGTGTCTGTAGCTTTGCCCCCGCTGCTTTCTCTCC
TTTGTCCCCGTCTGTCTTCTGTCTCAGGCCCCGCTGCTGGGTCCCCTTCTGTCTTTGCGTGGTTCTTCTGTCTTG
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGACGTCGGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGGTCTCTGCCCCGTGAGTGTCTGGAGCACCAGTGCCAGCGTTC
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCTGGGTTGAGCCTGGAAAAACCCAGGCATGTGCGGGTCTGGTGGCT
COCGGTGTGAGTTTGAATCGCGCAAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGCGGGGGAGTGTCTG
CTTCTCCCTTCTGCTTGGGAACAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCAAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCAGCCTGTAATCCAGCACTTTGGGAGGCCAAGCGGG
TGGATCACGAGGTGAGGAGTCTGAGACCATCTGGCCAACATGATGAAACCCCATCTGTACTAAAAACAAAAATTAGC
35 TGGGCGTGGTGGCGGGTGCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGCAGTGAGCCGACATTGCACCACTGCACTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA
AAAAAAAAAAATTCTAGTAGCCACATTAAAAAGTAAAAAGAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATTTTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTGAGC
TTTGTCTGCGGGATCCCGTGTGTAGGTCCCGTGGTGGCCATCTCGGCTGGACCTGCTGGGCTTCCCATGGCCATGGCT
40 GTTGTACCAGATGGTGCAGGTCCGGATGAGGTGCCAGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGGATGGTGCACG
TCTGGGATGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTCTCCAG

5
10
15
20
25
30

GGCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCTGTGAGCTGGATG
TGTGGTGTCTGGATGGTGCAGGTGAGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA
GGTCCGGGGTGAGGTCCGAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTACCC
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGAGACGGTCCGAGACCATGC
GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT
GTCCGGATGCTGCAGGTCCGGTGTGAGGTACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT
GGGGTGAAGGTCCGAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGAG
GCCCTCGGTGAGCTGGATGTGCAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCCGAGACCCCTGCGGTGAGCTGGATG
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGAGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGC
GGTCCGGGGTGAGGTCCGAGACCCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGAGGTCCG
AGACCCCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTGAGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCTGTGAACTGGATGTGCGGCGTCTGGATGGT
GCAGGTCTGGGGTGTGGTCCGAGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCC
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCGTCTGGATGGTGCAGGTCTGGGGTGTGGTCCGAGGCCCTCGGTGAGCTG
GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG
GTGCAGTCCGGGGTGAGGTCCGAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT
CACCAGGCCCTGCGGTGAGCTGGTGTGCGGTGTCCGGTGTCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATGTGCGGTGTCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCCGCTAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGAGGCCCTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG
GGGTGAGGTCCGAGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGAGGTCCGAGGCCCT
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCCTCGGTGAGCTGGATGTGGG
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT
CCGGGGTGAGGTACAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCCGAGG
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCAGTGAGCTGGATG
TGCTGTATCCGGATGGTGCAGGTCTGCGGTGAGGTCCGAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGC
GGTCCGGGGTGAGGTACAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCCG
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCGGTGAGCTGG
ATGTGCTGTATCCGGATGGTGCAGGTCTGCGGTGAGGTCCGAGGCCCTGCGGTGAGCTGGATGTGCAGTGTACGGATGG
TGCAGGTCCGGGGTGAGGTCCGAGGCCCTGCGGTGGGCTGTATGTGTGTCTGGATGGTGCAGGTCCGGGGTGAGTT
CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATATGCGGTGTCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCCGAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGAGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG
GGGTGAGGTACAGGCCCTCGGTGATCTGGATGTGGCATGTCTTCTCGTTAAG

35 Intron 3 (SEQ ID NO 6)

40

GTACTGTATCCCCACGCCAGGCCTCTGCTTCTCGAAGTCTTGAACACCAGCCGGCCCTCAGCATGCGCCTGTCTCCACT
TGCCTGTGCTTCCCTGGCTGTGCAGCTCTGGGCTGGGAGCCAGGGGCCCCGTACAGGCCTGGTCCAAGTGGATTCTGTG
CAAGGCTCTGACTGCCTGGAGCTCACGTTCTCTTACTTGTAAAATCAGGAGTTGTGCCAAGTGGTCTCTAGGGTTTGT
AAGCAGAAGGGATTTAAATTAGATGGAAACACTACCACTAGCCTCCTTGCCTTCCCTGGGATGTGGGTCTGATTCTCTC
TCTCTTTTCTTTTCTTTTGGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCCTGGGTTTAAGCGATTACACGAGCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTTAGGTCATGAGAGGATAAAATCCCACCCACTTGGCGACTCACTG
5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTTGAATG
GCTGTGAGATTTTGTCTGCAATGTTCCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTTCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC
TGA CTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCCTGCACACTCGAGTCCCTGGGGGGCTTGTGAC
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
10 GACCCCGACGTGGTGTCTGGGGCCATTTCTTGCATCTGGGGAGGGTCAGGGCTTCCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGAAAGAAT
TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG
CCTGGGGGGCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCTTTTCTACTCTGCTGGGCCTGCGGCTGCGGTCT
AGGGCACCAGCTCCGGAGCACCCGCGGCCCCAGTGCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT
15 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGCTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCCTCTCCATCTGAAGGGATGTGGCT
CTTTCTACCTGGGGGTCTGCCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG
AGGGGCATGGGTTACGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTACCCTGGGG
GTTGACCGCCGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

GTGGCTGTGCTTTGGTTTAACTTCCTTTTTAAACAGAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
25 CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCAACCCATTTGTGCGCACAGTGAGGTGGCCGAGG
TGCCGGTGCCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCTGGATCCGTGTCTGCTGTGGTGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCGGGACCAGGCCACGACTGCCAGGAGCCCACGGGCTCTGAGGATCCTGGACCTTGCCCCACGG
30 CTCCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCCGCCAG

Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCCTGGGTGGGCCGAGGGAGTGAGGTGACCCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCGGTGCCCTGCACCCAGTCTGAGCCAG
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCGCCCTGCCCTTGGGGTCTGGAGTGGTGGGGTTCAGAGAGAG

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTGGTCACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGG
ATGGCGGTCTGTGGGGTCTGATGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGA CTGTGGATGGCGGTCGTG
5 GGGTCTGATGTGGTGA CTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATG
TGGTGA CTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CT
GTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGG
CGGTCGTGGGGTCTGATGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGTGA TCGGTCA
CAGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGTGA TCGGTCA CAG
10 GGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTTGGTCCCGGGG
TCTGATGTGTGGTGA CTGTGGATGGCGATCGGTACAGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCT
GATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGGT
GA CTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGAT
GGCGGTTGGTCCCGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGA CTGTGGATGGCAG
15 TCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGG
TCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGT
GGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGTGA TCGGTACAGGGGTCTGATGTGTGGT
GA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGA CTGTGGAT
GGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTAGGGTCTGATGTGTGGTGA CTGTGGATGGCAGTCG
20 GTCACAGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGG
GGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGA CTGTGGATGGCGGTCGTGGGGTCTGAT
GTGGTGA CTGTGGATGGTGA TCGGTACAGGGGTCTGATGTGTGGTAGCTGCAGGTGGAGTCCAGGTGTGTCTGTAGCT
ACTTTGCGTCTCTCGGCCCGGCCCGTTTCCCAAACAGAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG
GGCTTGGCCGACAGTCCACACGTCTGATCGGAAGAAACAAGTGCCAGCTCTGGCCGGGGCAGGCCACATTTGTGGCTC
25 ATGCCCTCTCTCTGCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGGCAGTGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTAGGCACCTGCTCACGTTTGACTGCGCGGCCTCTCTCCAGTT
CCGCACTGCCCTTTGTTTCATGATTTGCTAAATGTCTTCTCTGCCAGTTTGTATCTTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAGCCGTGCTCTGCCAGCTGGCCCCCTCAGTGCTGGGTCTGAGGCCAAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGGCCGTGTTTGAGCCACGCCCGCTGAGCGGGCCTCTCAGTGCTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTGCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTCAGATGCCCTGTTAGCACTTGCTCGGC
35 TCTAGGGGACACTCGTGTCCACGCATGAGGCTCAGAGACCTCTGGGCGAATTTCTTGGCTCCAGGGTGGGGGTGGAG
GTGGCCTGGGCTGCTGGGACCCAGACCCTGTGCCCGGACGTGGGCAGCAACTCCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCACAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCATGTGTGTCTGCAGAGACTCGGCCCGGCCAGCCACGATGGCCCTGCATTCAGCCCAGCCCCGCACTTCATCACA
AACACTGACCCCAAAGGGACGGAGGGTCTTGGCCACGTGGTCTGCTCTCAGCACCCACCGGCTCACTCCCATGTG
40 TCTCCCGTCTGCTTTCGAC

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCCC
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC
5 TCCCCAGGCCCCCGAAACATGGCTCGGCTTGCGGCAGCCGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCCGATACAAAA
GGATTTTATCCGATTCTCATTCTGTCCCTGTGCTGTGACCCCGCGAGGGCGGGGCTCTTCTCTGTGACTAGATTT
CCCATCTGGAAGTGGGGGTTGACCGTGTAGTTGCTCCTCTCGGGGGGCTGTGGTGGCCATGGGGCAGGCGGCCTGG
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCAGGTGGTAGAGCCACAGTGCTGTGCCACATCACGCTCT
CTGGATTTTAAGTAAAACACACACCTCCCGGAGGCATCTGCCTGCGACCTGTGTGTGCTGGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTCATCAGCAAGGTCATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATATAACTATTAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATATTTATTAAAGTATAATTAGAAATATTAAGTAGTACACAGTTCTGAAAAA
15 CACAAATTGCACATGGCAGCAGAGTGAATTTTGGCCGAGGGACAGTGTGCACATGTGTGTAAGCGGCCCCCAGGCCCCAC
AGAATTCGCTGACAAAGTCACCTCCCCAGAGAAGCCACCACGGGCCTCCTTCGTGGTGTGTAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCCTCATGCCCTGACAGACAGGA
GGTGAAGTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCCGAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAACGTCCTTCAAAACCTGTTGCCCCAAAACTAAGAACAGAGAGAGTTTCCCATCCCATGTGCTCACAGGGGCGTA
20 TCTGCTTGCCTTGACTCGCTGGGCTGGCCGGACTCTAGAGTTGGTGCCTGTGCTTCTGTGCAAAAAGTGAGTCTCTCT
GCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTCTTTCTTTCTTTTCTTTTGTGAGACGGAACGTCA
CTGTTGTCTGCCTGGGCTTGAGTGACGTGGCGGATCTCAACTCACTGCAACCTCCGCCTCCCGGGTCCAGCATTTCTC
CTGCCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCCTGGCTAATTTTGTATTTTGTAGAGAG
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGAAAGCCTCTTTTTAAGGTGACCACCTATAGCGCTTCCCGAAAATAAC
AGGTCTTGTTTTTGCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTGTCTGAAAACGCA
CCCTTGGCATCCTTGTTTGGAGAGTTTCTGCTTCTCGTTGGTCACTGTAAGTCTAGGGGCAAGGTTGTATCCGTTGGCGC
GCAGCGGCTACATGTAGGGTCACTGAGTCTTTACCGTGGACAAATTCCTGAAAAAAGGAGTCCGGTTAAGCAT
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTTAAGAAACCTTAATGAAAGAAAACCTTGATGATTC
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTCAGCCGCCCCAGTGCATGGTGAGAGTGGGGAGCAGGGATTG
TTGTTTCAAGAGTCTCATCTGGTATGTTTCTGAGGTGTTTCCGGGCTGAATGGTAGAGTGTCTGTTGTGTATGAGGT
TCTGTGTCTGTGTGGCTCGGTTTGTAGGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC
CCCAG

35

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTTGATTGCTTTTGATGCATTAGTGTTAATATTCCTGGTGC
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACCAGACAAGGTTGAGCCCCCTTCTTGGTATGAAGCCGCACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTTCTGTACGTCACCCAGGTTCC

CGTTAGGGTCCCTTGGGGAGATGGGGCTGGTGAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTGCCCATCCCCTTGATGGGGTCTACACCAAGGACGCACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACCTCATCTTTACA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCTG
TGTCCAAGTGTCTCATTGTTTCTGTTCCACCTGTGAGTGAGAACATGTGGTGTGGTGTCTTCTCCTTGCAATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTGGTGTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAATGGTTGAA
CTAGTTTACACTCCCAACAGTGTAAGTGTCTGGTGCTGGAGAGGATGTGGACAGCAGTTATTTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC
TCCGGGTTTTCTGTGCATCTTTTGAAACTTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAACAAAATTCTTGTAACACAACTTGCTCTGGGATTTGGA
15 GGAAAGTGTCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCGCTGAGGGTCACACAGTGACCATGCCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGCGAGGTTCCAGCCCCAGCTTCTTACCCTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCGCGCCTGATGGCCTTCGTTCTTCAGCTGGCACAGAATTGCACAA
GCTGATGGTAAACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGA
20 GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAAATTGTTGACATTAGTCCC
TCGTAGACAGATACTACGTAAAAAGTGTAAGTTAACCTTGCTGTGTATTTTCCCTTATTTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCGCGTGCCGTGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCTGG
25 CACCGCAGCGTTGTCTCTGCAAGTCTCTCTCTGCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCCTGGGGGCGCAGGGGCACCTTCGGGAGGAGTGGGTACCGTGACGGCCCTGGTCTGCAGAGACGACCCAGGTT
ACACACGTGGTGAGTGACGGCGGTGACCTGGCTCCTGTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGCACAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCCAGGGTGACCTGAGCCTGCGGA
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTCGCTGCGGTACGTTCTGCGTGGGGTGTGTTGGGATCGGTGGG
30 AGAATTTGGATTTGCTGAGTGCTGTCTTGAACACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTTGTGAAT
CAAATAAAATCAGGCACAGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGTTGGGCTGTGTTTAAAGTGCATTTGACGAGGGACGAGAACTTGAAAGCTGTAAAGGGAACCTCAGAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTGTGAAAACCCATTGGACCCGCCCTCAAGTCCACCC
TCCAGGTCCACCCTCCAGGGCGCCCTGGGCTGGGGGTATGCTTGGCGTTCTTGTGCGCAGCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGAACGCTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGCGCAGGGAGGTGGCTCAGA
GTGTATGTTGGGGTCCCACGGGGGAGAACTCTGTCTCTGATGAGTCCGCAGCCATGTAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGGAGGGGGACGCC
GGGGCCACAGCAGAGGCCGAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCAGGGGGGCTCCCTG
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCAGAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCTTAAAAGAAGGTGGGATGGTGGCAATTTCTTGTCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCTCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGTCATGCTCACCTACCTGTCTGCCC
5 GGGAGACAGGGAAGCACCCTGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT
GCTCCAAATCACCACTTCTCTGGGGTTTTCCAAAGCATTAAACAAGGGTGTGAGGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGAAGTGAGCCTGTGCCCCGGCTGGGGCAGGTGTGCTGTCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGGCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGTCTCAACGGGAGCAGTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAAACCCGAACACAGGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCCTGGGGCCCTGCTGGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTTGTTTTCCCAGAGCCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTTCTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGACACGCCTGTAGTCCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAGTATCAGCATTCCAAAACCATAGTGACAGGTGTTTTTTTATTCT
TGTCCTTCGATAATATTTACTGGTGCTGTGCTAGAGGCCGAACTGGGGGTGCCTTCCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAACTGGGGTCTGTGCTTCTGAGTTAACAGTCCAGATC
25 TGGACTTTGCCTCTTTCCAGAAATGCTCCCTGGGGTTTGCTTCATGGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGGCCAGCATGTCCCTG
TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCTCCACTGGCTTTGTCTGCATGATTCCACATTTCTGGGCTCCCAGCACCTCTTCGCCTCTCCAGGCACCTCT
GCAGTGCTGGCCATACCAGTCAGCTGTGAAGTGTCCACTGCTTATTTTGCTCCCCATGAAATGTATTTTTTAGGACAGGC
30 ACCCTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGCAGGATGGGTGGGCATCAGGTTCATCAGATGTGGTCCAATG
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGACAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAAGTCACTCCTCTGTGCTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAAAGGAAAAGTGGTAAGATGGGAATTTTCTTGTCAGATTTTAGTCT
35 TCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACAATAGAACAACCGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTGTAATTTTTTTTCTGAGAAAAC
GACTGGAAGCAAATAAGTTGTGTCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAAACAAC
ACCAGCAACAGAAATAAAACAAAAGACTCAAAGGGAAGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAACTCAGCTTGCTGAGCCACAGTGAAA
40 ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGCACATTATCCTCTCACTTT

GTTCTCCTAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCGAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
5. GAGACCTGTCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTGTCATGGAAGTCCTCACAATGTCTGTCTTCCCAGTAATCCACTTCTGAAGTGA
CCAGACATTATCACGGGTCTTATTTACCATTTCCAGTGTTCCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAAACTTGCTCTGCCATTAAACATTTTTCAAAGAATTTTGAAGAAT
GTTTAATGGCACAAAACGTTTATTTCAATGTAGCAGTGTTCAAAGCTGGATGTAAAAGAACACACCCAGGAGCCTGCCG
10. TGAATGTCATGTGTGTTTCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCAGTGCACCTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCAACTCAGTGTTCTCCACAAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTCTGGTG
15. GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCAGCCCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTGCAGGAGGGGGCTGGGTGT
GGGCGAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20. GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT
GCCTGCAAGGCTGATGGTGA CTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA
TGATGTGTGTTTCGTGCACAGTCGTGTGGGCATTACAGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCTCGTGTTACCCCCGCTAGGTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
25. GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTTCGAGCTG
CCGGGTGAGGGCCAGGCCGATTCTACTGGGAAGAGGGATAGTTTCTGTCAAATGTTCTCTTTCTGTTCCATCTGA
30. ATGGATGATAAAGCAAAAAGTAAAACTTAAATCCAGAGAGGTTTCTACCGTTTCTCACTCTTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAGTGGTGCTGCTGCCTGTGCACAGTTCTGTTTCGCGTG
35. GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT
GCACTGGCCGTGGGACGTATGGAGGCCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGCTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGTCCCCACACAGCCCGGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCTGTCTGGCTGGTCAGGGGTGCCCTGCCAAGAATCG
40. ACAACTTTATCACAGAGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTATTACCCAGGGCCG
AGGCTGCGCGAATTACCGTGCACACTTGATGTGAAATGAGGTCGTCTATCGTGGAAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAAAATGGTTTTTAACCCGAGTGCTTGCCTTCATGCTCTGGCAGGGAGGGC
5 AGAGCCACAGCTGCATGTTACCGCCTTTGACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTCTGCCTTCTCTCTCTGCCTCAAATCTTCCCTCGTTTGATCTCCCTGACGCGTGCCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGTCTGGGGCCTCCTTGGGCCATGATGAGGTGAGGAGTGTTC
CAGGTGAAAACCTCCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTC
10 ATTTCCCCACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGCCCCGTGCTGGCCGCGCTCCACACGGGCTTGGGGTGGACGCCCCGACCTCTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCCCTACCCATGCTAGGTGTTTCCCTCCTGGGTGAGAGCGTGGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCCTGGCTTCCGTTGTGCTAAATGGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTGCTGGCTTACTGCTGTGATCTCAGGTCAATCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGGCTCAGGCACTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGAGGGTCAGACCAGGTAC
20 ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAAAACAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAAGAAGACTGACAAATGCAGTTTCTTGAAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTCTGAGATGAGATGATGGGTCTCACACCATCACCCAGACCCAGGGTTTATG
CACCAAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGTTGTCTGACGAAGGGCAG
GATTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAAACCTTAGAGGCCTTCCCGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGCAGTGTACCTGTGCACACACAGACACGCAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCCATGAGGAAACCATGCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCAT
GCCACACCCACGAGACCGTCTGATTAGGAGGCCTTTCCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCAACCACTGTGAGGCTCGTCTGCCCCCCTCTCGT
GGGGTGAAGCAGACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCCTGGGGAA
GCGCTGGGGGGCCTGGTCTCTCTGTTTGGCCATGGTGGGATTGGGGGGCCTGGCCTCTCTGTTTGGCCTGTGGTGG
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCGTATCACACGACAGAGCCCCGCGCGTCTCTGCTTCCAGTCAACG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCCGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTTCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTTCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCAGGGCCCCACCTGCCAGGGGTATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGCTGCTGAAGTGAGACGCCCCCGGGCCTGACCCTGGGGGCTGGAGCCAGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCCGCCCCGCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCCGGTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACCAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACACGTCTTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

10

15

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCAGGGCCCATGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCAGGGCCCATG
GCCTGGCTGGGCCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGGCGCGGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGGCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGTGGTCTGGGTGG
GGGGGAAGATGGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGGCCATCAGCTTTCATGGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCAATTCGCCCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTATGCTCTCTC
ATCCTCTTATCATCTCCCAGTCTCATCTCTCATCCTCTTATCATCTCCCAGTCTCATCTGTCTTCTCTTATCTCCCAGT
CTCATCTGTCTATCCTCTTACCATCTCCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCTTCTCAGGCAGAAGGAAGTGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGAGTCTTGGGGTGAAGAAACAGCCCCCTCCTCAGAAGTTGGCTTGGGCCACACGAAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGTAGTGAGCCCGGACGTGCCTGGTGTGCGGGTGGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGGCCCAAGTCCACAGACTG
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTTGGCAGCCGGCCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGTGGGGCGGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCAGACCAACaGGTCAGGCCATT
GTTCACTATCCATCTTCTACAAAGCTCCAGATTCTGTTTCTCCGGGTGTTTTTGTGTAATTTTACTCAGGATTACT
TATATTTTTTGCTAAAGTATTAGACCCCTTAAAAAAGGTATTTGCTTTGATATGGCTTAAGTCACTAAGCACCTACTTTAT
TTGCTGTTTTTATTTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCCAGGTTGTTAGTGAGTGGCAC
AGTCATGGCTCGCTGTAGCCGCAACCCCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCCAGAGTGCTGGGATTACAG
GTGTGAGCCACTGCCCTTGCTGGCACTTTTAAAAAACCCTATGTAAGGTGAGGTCCAGTGCGCTTCCACACCTGTATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAAGACCAGCATGGGTAACATAGGGAGACCCC

20

25

30

35

40

ATCTCTACAAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCTGTCTCAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG
5 GAAGAAGGAAGAAAGAAGGAGAAGGAGGCCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAAATAACA
AAGTTTTAAAGGGAAAGAAAAACCCAGCTCTTTGGAATTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGCACCTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
GCACTTGTTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTATCTGGAATTATCTGG
TGGGCCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCTGAGGGCACACGGCCCTGCCACGCTCGATTTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTTCAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAATGGAATAG
15 CAGTACAGGGAAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCCCCCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

20

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Spl-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA_nTTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losten Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei
10 anderen TATA-losten Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer
30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmen aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIAquick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄ pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-
plattieren.

- Für die Transfektion wurden pro Well jeweils 2×10^5 HEK 293 Zellen in einer 24-
well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV
β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-
freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer
Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die
HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24
Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität
geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8]
mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15
Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O;
2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM
Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht
gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-
Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- 35 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

- Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.
- 5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.
- Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.
- 10 Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.
- 15 Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen
Telomerase-Untereinheit.
5
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser
Sequenzen handelt.
10
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
15
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß
20 es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide
oder Proteine kodieren.
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
- 25 7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-
toren gemäß einem der Ansprüche 4 bis 6.
30

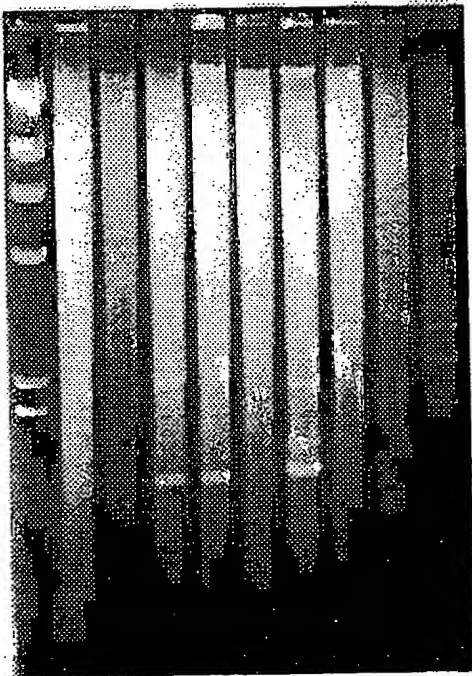
9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screenet.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reporter-genkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

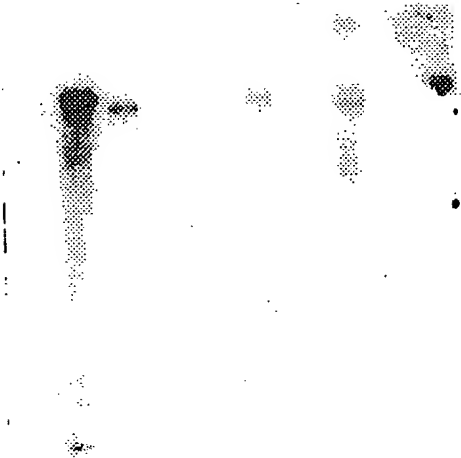


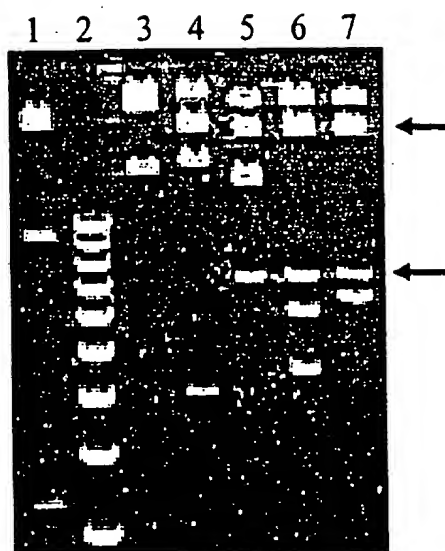
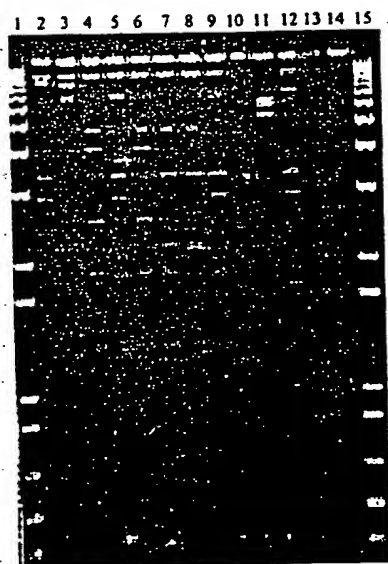
Fig. 2

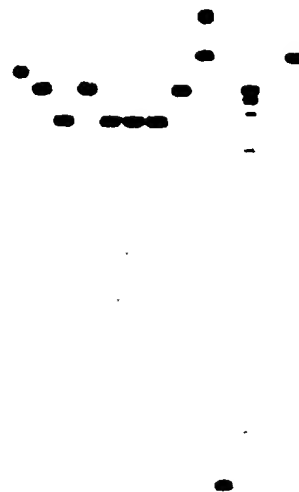
Fig. 3

A



B

2 3 4 5 6 7 8 9 10 11 12 13 14



GAGCTCTGAA	CCGTGGAAAC	GAACATGACC	CTTGCCCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	70
GTGGTGTGCA	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCCTC	CATCATTATT	140
CATCTTTCACC	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	210
TACAAACACC	ACTCTTTTAC	TAGGCCACACA	GAGCACGGGC	CACACCCCTG	ATATATTAAG	AGTCCAGGAG	280
AGATGAGGCT	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTCC	TCTAGACTAG	350
TAGACCCTGG	CAGGCACTCC	CCCAAATTCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	420
GAGACTCAGC	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCTCCAG	GCCTCAGCTT	490
CTCCAGCAGC	TTCCTAAACC	CTGGGTGGGC	CGTGTTCCAG	CGCTACTGTC	TCACCTGTCC	CACTGTGTCT	560
TGCTCAGCG	ACGTAGCTCG	CACGGTTCCT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	630
GCGTTGAAGG	GAGGAGATT	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	700
GATGCAGGTT	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TCCAGGCGC	TCCCCGTCTC	CTGTCATCTG	770
CCGGGGCCTG	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	840
CTAGGGTCTC	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	910
CATTTGGGTG	TGAAAGTAGG	AGTGCCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	980
CCGCCAGGGA	CCCGCCCTTC	TCTGCCCAGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	1050
TTCCACAAGC	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	1120
CCTGGGAATT	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCCG	CTGTTTTATT	1190
TTAATAGCTA	CAAAGCAGGG	AAATCCCTGC	TAAAATGTCC	TTTAAACAAAC	TGGTTAAACA	AACGGGTCCA	1260
TCCGCACGGT	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	1330
GAATTACGCT	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTC	1400
ACCCCATGGC	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	1470
CCTTTTACTA	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	1540
TGGGGATGGG	GGAAACCGGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	1610
AGAGATGCCC	ACGTCTTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCG	GCCCCAGGGC	CTTTGACGCT	1680
GTGATCTCCG	TGAGGACCCT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCGGAAAA	GTATATCCAGG	1750
GGTTCGTGGA	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGGAGCA	GTACGTCTGA	1820
GGCTGAAAAA	GGAGGGAGGG	CCTCGAGCCC	AGGCCCTCAA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	1890
GGGACCCCTC	ACGGAGCCTG	CAGCAGGAA	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	1960
TCCGGCTCTC	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAAA	2030
GCAACAGGAA	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	2100
CAGGGCTGAA	GTGCCTCCGG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	2170
TACTTACTTT	CTGAGACAGA	GTTATGCTCT	TGTTGCCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	2240
CTGCAACCTC	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTCAGGC	2310
GTGCACCACC	ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	2380
TCTCAAAATC	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	2450
CACTGCACCT	GGCCTATTTA	ACCATTTTAA	AACCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	2520
CATGGAGTTC	AATTTCCCCT	TTACTCAGGA	GTTACCCTCC	TTTGATATTT	TCTGTAATTC	TTCTGAGACT	2590
GGGGATACAC	CGTCTCTTGA	CATATTCACA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	2660
GGGCAGCTGG	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	2730
ATCAGGGCGC	AAGTGTGGAC	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	2800
AAGTCCATCC	CTCCTACTCT	ACTGGGATTG	AGCCCCCTCC	CTATCCCCCC	CCAGGGGCAG	AGGAGTTCCT	2870
CTCACTCCTG	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACTGCTGGT	ACTGAATCCA	CTGTTTCATT	2940
TGTTGGTTTG	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	3010
CGCGATCTTG	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	3080
GCTGGGATTA	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGATTTTT	TAGTAGAGAC	GGGGGTGGGT	3150
GGGGTTACCC	ATGTTGGCCA	GGCTGGTCTC	GAACCTTCTG	CCTCAGATGA	TCCACCTGCC	TCTGCCTCCT	3220
AAAGTGCTGG	GATTACAGGT	GTGAGCCACC	ATGCCCAGCT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	3290
TCTGAGGTAG	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAATT	TTTTTATTGT	3360
TGTTAGAACA	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	3430
TGCACCCATA	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA			

Fig. 4 (Fortsetzung)

CACAGCCTAG GCCGATTGGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126

6 / 15

Fig. 5

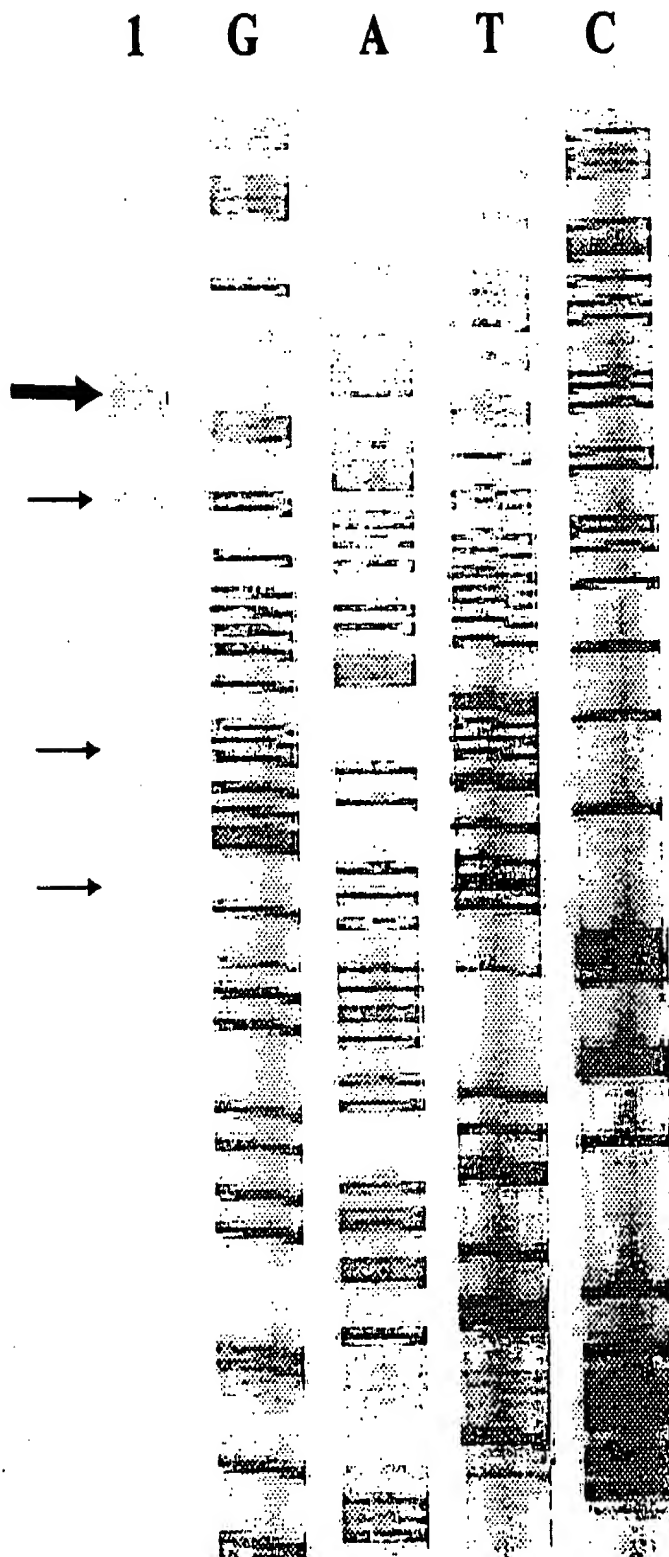


Fig. 6

GTTCAGGCA	GCGCTGCGTC	CTGCTGCGCA	CGTGGGAAGC	CCTGGCCCCG	GCCACCCCCG	CGATGCCGCG	70
CGCTCCCCGC	TGCCGAGCCG	TGCGCTCCCT	GCTGCGCAGC	CACTACCGCG	AGGTGCTGCC	GCTGGCCACG	140
TTCTGTGCGG	GCCTGGGGCC	CCAGGGCTGG	CGGCTGGTGC	AGCGCGGGGA	CCCGGCGGCT	TTCCGCGCGC	210
TGGTGGCCCA	GTGCTGGTGG	TGCGTGCCCT	GGGACGCACG	GCCGCCCCCC	GCCGCCCCCT	CCTTCCGCCA	280
GGTGTCTCTG	CTGAAGGAGC	TGGTGGCCCC	AGTGCTGCAG	AGGCTGTGCG	AGCGCGGCGC	GAAGAAGCTG	350
CTGGCCCTTC	GCTTCGCGCT	GCTGGACGGG	GCCCCGCGGG	GCCCCCCCCA	GGCCTTCACC	ACCAGCGTGC	420
GCAGCTACCT	GCCCCAACAC	GTGACCGACG	CACTGCGGGG	GAGCGGGGCG	TGGGGGCTGC	TGCTGCGCCG	490
CGTGGGCGAC	GACGTGCTGG	TTCACTTGCT	GGCAGCTGTC	GCGCTCTTTG	TGCTGGTGGC	TCCCAGCTGC	560
GCCTACCAGG	TGTGCGGGCC	GCCGCTGTAC	CAGCTCGGCG	CTGCCACTCA	GGCCCCGGCC	CCGCCACACG	630
CTAGTGGACC	CCGAAGGCGT	CTGGGATGCG	AACGGGCGCT	GAACCATAGC	GTCAGGGAGG	CCGGGGTCCC	700
CCTGGGCGTG	CCAGCCCCCG	GTGCGAGGAG	GCGCGGGGGC	AGTGCCAGCC	GAAGTCTGCC	GTTGCCCAAG	770
AGGCCCAGGC	GTGGCGCTGC	CCCTGAGCCG	GAGCGGACGC	CCGTTGGGCA	GGGGTCTCTG	GCCCCACCGG	840
GCAGGACGCG	TGGACCGAGT	GACCGTGGTT	TCTGTGTGGT	GTCACCTGCC	AGACCCGCGG	AAGAAGCCAC	910
CTCTTTGGAG	GGTGCCTCT	CTGGCAGCGG	CCACTCCCAC	CCATCCGTGG	GCCGCCAGCA	CCACGCGGGC	980
CCCCCATCCA	CATCGCGGCC	ACCACGTCCC	TGGGACACGC	CTTGTCCCCC	GGTGTACGCC	GAGACCAAGC	1050
ACTTCCTCTA	CTCCTCAGGC	GACAAGGAGC	AGCTGCGGCC	CTCCTTCCTA	CTCAGCTCTC	TGAGGCCCAG	1120
CCTGACTGGC	GCTCGGAGGC	TCGTGGAGAC	CATCTTTCTG	GGTTCAGGC	CCTGGATGCC	AGGGACTCCC	1190
CGCAGGTTGC	CCCGCCTGCC	CCAGCGCTAC	TGGCAAATGC	GGCCCCGTGT	TCTGGAGCTG	CTTGGGAACC	1260
ACGCGCAGTG	CCCCTACGGG	GTGCTCCTCA	AGACGCACTG	CCCGCTGCGA	GCTGCGGTCA	CCCCAGCAGC	1330
CGGTGTCTGT	GCCCCGGGGA	AGCCCCAGGG	CTCTGTGGCG	GCCCCCGAGG	AGGAGGACAC	AGACCCCCGT	1400
CGCCTGGTGC	AGCTGCTCCG	CCAGCACAGC	AGCCCCTGGC	AGGTGTACGG	CTTCGTGCGG	GCCTGCCTGC	1470
GCCGGCTGGT	GCCCCCAGGC	CTCTGGGGCT	CCAGGCACAA	CGAACGCCGC	TTCTCAGGA	ACACCAAGAA	1540
GTTTCATCTC	CTGGGGAAGC	ATGCCAAGCT	CTCGTGCAG	GAGCTGACGT	GGAAGATGAG	CGTGCGGGAC	1610
TGCGCTTGGC	TGCGCAGGAG	CCCAGGGGTT	GGCTGTGTTT	CGGCCGAGA	GCACCGTCTG	CGTGAGGAGA	1680
TCCTGGCCAA	GTTCTGTCAC	TGGCTGATGA	GTGTGTACGT	CGTCGAGCTG	CTCAGGTCTT	TCTTTTATGT	1750
CACGGAGACC	ACGTTTCAA	AGAACAGGCT	CTTTTCTAC	CGGAAGAGTG	TCTGGAGCAA	GTTGCAAAGC	1820
ATTGGAATCA	GACAGCACTT	GAAGAGGGTG	CAGCTGCGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCAGCATC	1890
GGGAAGCCAG	GCCCGCCCTG	CTGACGTCCA	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	1960
TGTGAACATG	GACTACGTCG	TGGGAGCCAG	AACGTTCCGC	AGAGAAAAGA	GGGCCGAGCG	TCTCACCTCG	2030
AGGGTGAAGG	CACGTGTCAG	CGTGCTCAAC	TACGAGCGGG	CGCGGCGCCC	CGGCCTCTCT	GGCGCCTCTG	2100
TGCTGGGCCT	GGACGATATC	CACAGGGCCT	GGCGCACCTT	CGTGCTGCGT	GTGCGGGCCC	AGGACCCGCC	2170
GCCTGAGCTG	TACTTTGTCA	AGGTGGATGT	GACGGGCGCG	TACGACACCA	TCCCCCAGGA	CAGGCTCAGC	2240
GAGGTCATCG	CCAGCATCAT	CAAACCCAG	AACACGTACT	GCGTGCGTCG	GTATGCCGTG	GTCCAGAAGG	2310
CCGCCCATGG	GCACGTCCGC	AAGGCCTTCA	AGAGCCACGT	CTCTACCTTG	ACAGACCTCC	AGCCGTACAT	2380
GCGACAGTTC	TGGGCTCACC	TGCAGGAGAC	CAGCCCGCTG	AGGGATGCCG	TCGTATCGA	GCAGAGCTCC	2450
TCCCTGAATG	AGGCCAGCAG	TGGCCTCTTC	GACGTCTTCC	TACGCTTCAT	GTGCCACCAC	GCCGTGCGCA	2520
TCAGGGGCAA	GTCTACGTC	CAGTGCCAGG	GGATCCCGCA	GGGTCCATC	CTCTCCACGC	TGCTCTGCAG	2590
CCTGTGCTAC	GGCGACATGG	AGAACAAGCT	GTTTGCGGGG	ATTGCGCGGG	ACGGGCTGCT	CCTGCGTTTG	2660
GTGGATGATT	TCTTGTGGT	GACACCTCAC	CTCACCCACG	CGAAACCTT	CCTCAGGACC	CTGGTCCGAG	2730
GTGTCCTTGA	GTATGGCTGC	GTGGTGAAGT	TGCGGAAGAC	AGTGGTGAAC	TTCCCTGTAG	AAGACGAGGC	2800
CCTGGGTGGC	ACGGCTTTTG	TTCAAGATGCC	GGCCACGGCG	CTATTCCCCT	GGTGCGGCCT	GCTGCTGGAT	2870
ACCCGGACCC	TGGAGGTGCA	GAGCGACTAC	TCCAGCTATG	CCCGGACCTC	CATCAGAGCC	AGTCTCACCT	2940
TCAACCGCGG	CTTCAAGGCT	GGGAGGAACA	TGCGTCGCAA	ACTCTTTGGG	GTCTTGCGGC	TGAAGTGTCA	3010
CAGCTGTGTT	CTGGATTGCG	AGGTGAACAG	CCTCCAGACG	GTGTGCACCA	ACATCTACAA	GATCCTCTCT	3080
CTGCAGGCGT	ACAGGTTTCA	CGCATGTGTG	CTGCAGCTCC	CATTTTCATCA	GCAAGTTTGG	AAGAACCCCA	3150
CATTTTTCCT	GCGCGTCATC	TCTGACACGG	CCTCCCTCTG	CTACTCCATC	CTGAAAGCCA	AGAACGCAGG	3220
GATGTCGCTG	GGGGCCAAGG	GCGCGCGCGG	CCCTCTGCCC	TCCGAGGCGG	TGCAGTGGCT	GTGCCACCAA	3290
GCATTCTGCT	TCAAGCTGAC	TGCACACCGT	GTCACCTACG	TGCCACTCCT	GGGGTCACTC	AGGACAGCCC	3360
AGACGCAGCT	GAGTCGGAAG	CTCCCGGGGA	CGACGCTGAC	TGCCCTGGAG	GCCGCAGCCA	ACCCGGCACT	3430
GCCCTCAGAC	TTCAAGACCA	TCCTGGACTG	ATGGCCACCC	GCCACAGCC	AGGCCGAGAG	CAGACACCAG	3500
CAGCCCTGTC	ACGCCGGGCT	CTACGTCCCA	GGGAGGGAGG	GGCGGGCCAC	ACCCAGGCCC	GCACCGCTGG	3570
GAGTCTGAGG	CCTGAGTGAG	TGTTTGGCCG	AGGCCTGCAT	GTCCGGCTGA	AGGCTGAGTG	TCCGGCTGAG	3640
GCCTGAGCGA	GTGTCCAGCC	AAGGGCTGAG	TGTCCAGCAC	ACCTGCCGTC	TTCATTCCC	CACAGGCTGG	3710
CGCTCGGCTC	CACCCAGGGG	CCAGCTTTTC	CTCACCAGGA	GCCCGGCTTC	CACTCCCCAC	ATAGGAATAG	3780

Fig. 7

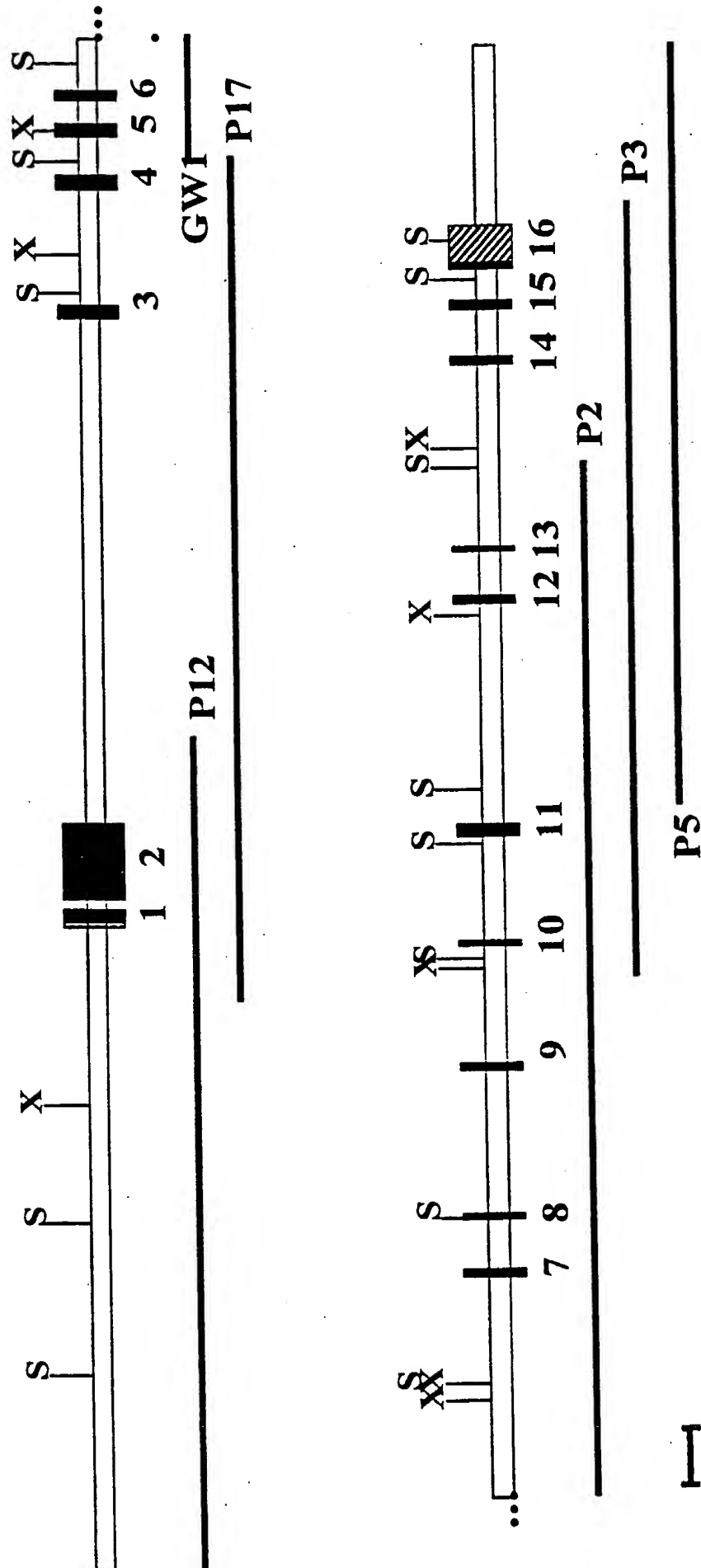


Fig. 8A

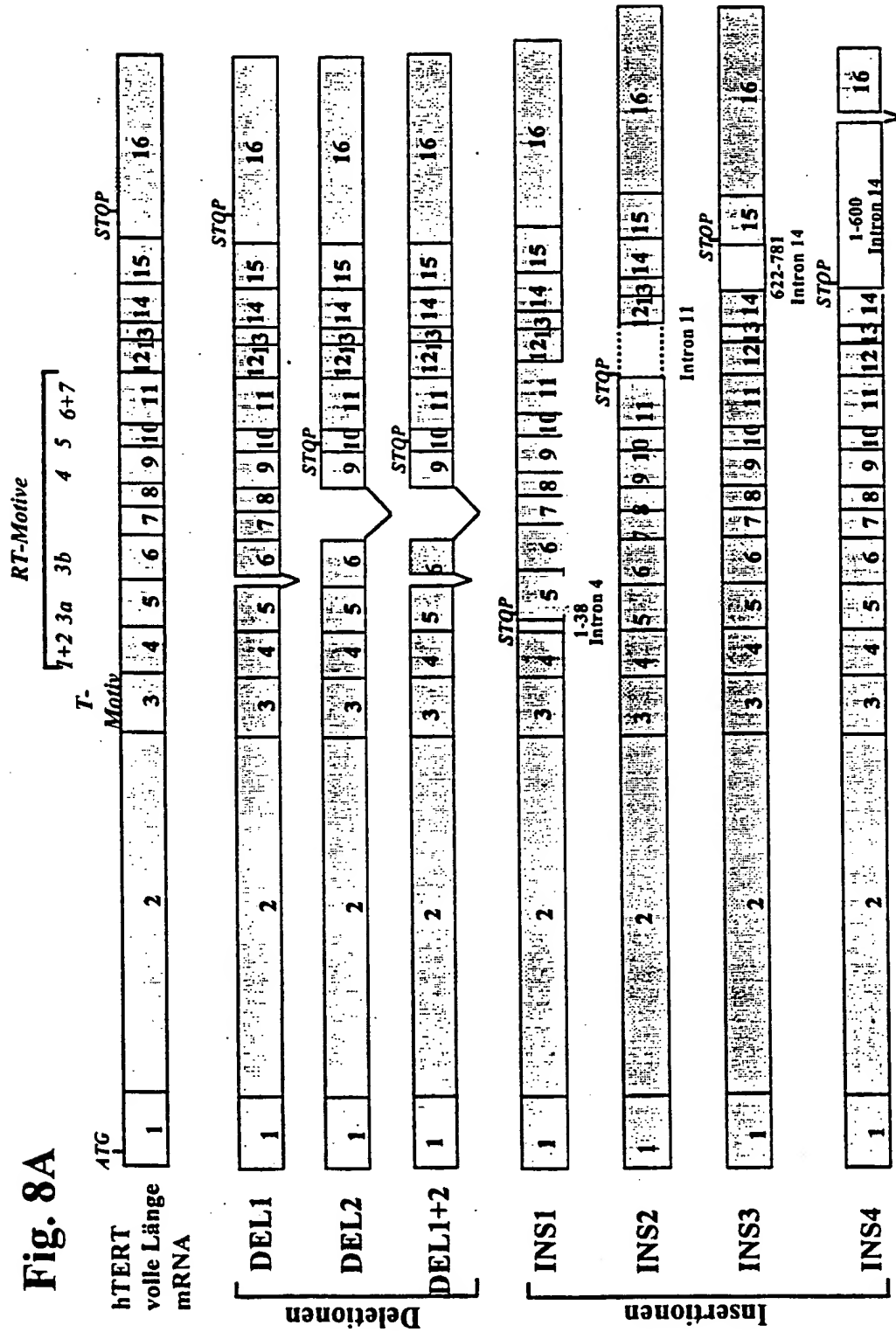


Fig. 8B

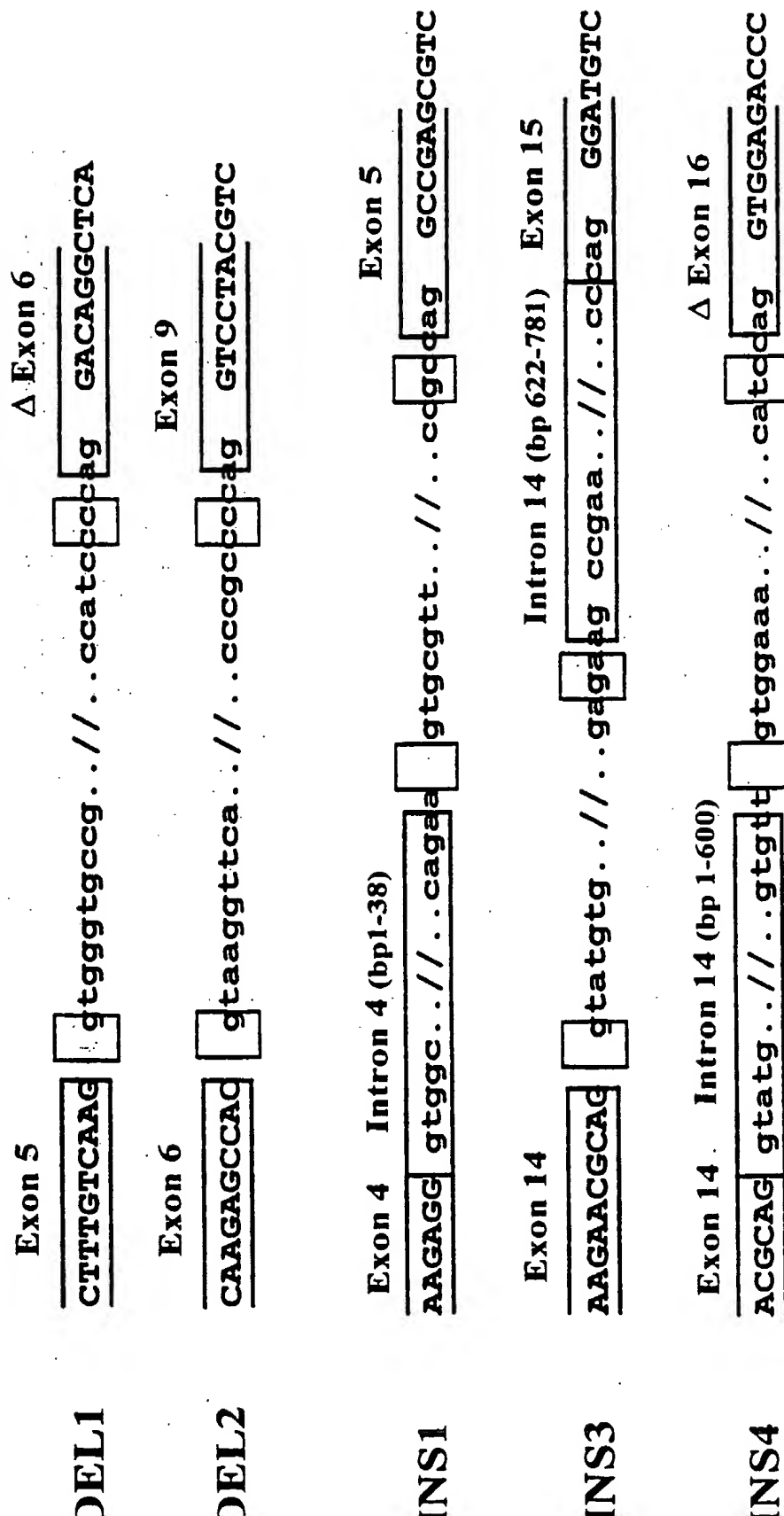


Fig. 9

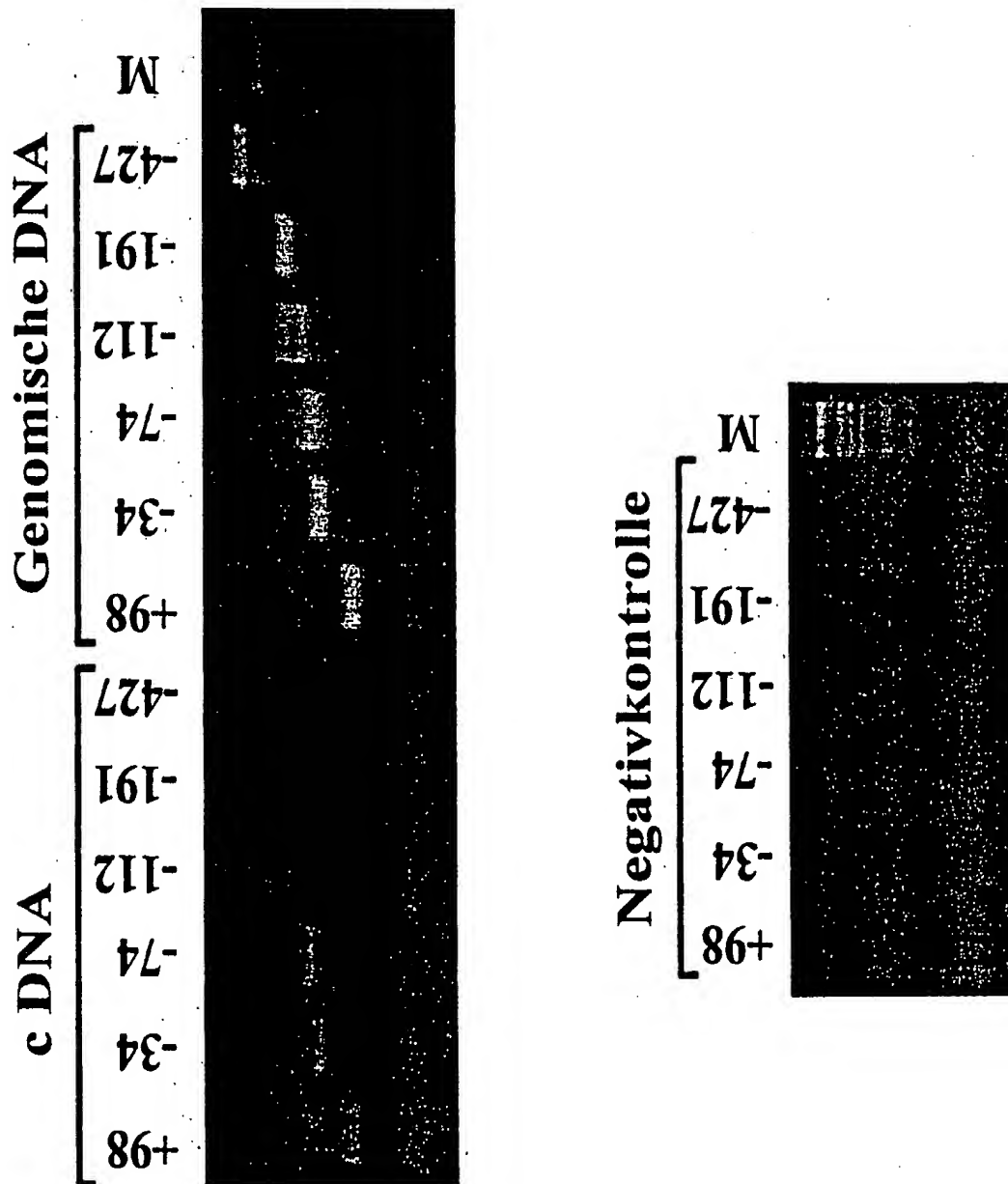


Fig. 10

ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	-11204
ATGAGACCCT	GTCTCAAAAA	AAAAAAGGAA	AATTGAAATA	ATATAAAGCA	TCTTCTCTGG	CCACAGTGGA	-11134
ACAAAACCAG	AAATCAACAA	CAAGAGGAAT	TTTGAAAAC	ATACAAACAC	ATGAAAAATTA	AACAATATAC	-11064
TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TTAAAAAGGA	AATTGAAAAA	TTTATTTAAG	CAATGATATA	-10994
CGGAAACATA	ACCTCTCAAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	-10924
AGCAGCTACA	TCAAAAAAGT	AGAAAAGCCA	GGCGCAGTGG	CTCATGCCTG	TAATCCCAGC	ACTTTGGGAG	-10854
GCCAAGGCGG	GCAGATCGCC	TGAGGTCAGG	AGTTTCGAGAC	CAGCCTGACC	AACACAGAGA	AACCTTGTCG	-10784
CTACTAAAAA	TACAAAAATTA	GCTGGGCATG	GTGGCACATG	CCTGTAATCC	CAGCTACTCG	GGAGGCTGAG	-10714
GCAGGATAAC	CGCTTGAACC	CAGGAGGTGG	AGGTTGCGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	-10644
TGGGTAACAA	GAGTGAAACC	CTGTCTCAAG	AAAAAAGAAA	AAGTAGAAAA	ACTTAAAAAT	ACAACTAAT	-10574
GATGCACCTT	AAAGAACTAG	AAAAGCAAGA	GCAAACTAAA	CCTAAAATTG	GTAAAAGAAA	AGAAATAATA	-10504
AAGATCAGAG	CAGAAATAAA	TGAAACTGAA	AGATAACAAT	ACAAAAGATC	AACAAAATTA	AAAGTTGGTT	-10434
TTTTGAAAAG	ATAAACAAAA	TTGACAAACC	TTTGCCCAAG	CTAAGAAAAA	AGGAAAGAAG	ACCTAAATAA	-10364
ATAAAGTCAG	AGATGAAAAA	AGAGACATTA	CAACTGATAC	CACAGAAATT	CAAAGGATCA	CTAGAGGCTA	-10294
CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACTTAGAA	AAATCTAGATA	AATCCCTAGA	TGCATACAAC	-10224
CTACCAAGAT	TGAACCATGA	AGAAATCCAA	AGCCCAAAACA	GACCAATAAC	AATAATGGGA	TTAAAGCCAT	-10154
AATAAAAAAGT	CTCCTAGCAA	AGAGAAGCCC	AGGACCCAAT	GGCTTCCCTG	CTGGATTTTA	CCAATCATTT	-10084
AAAGAAGAAT	GAATTCCAAT	CCTACTCAAA	CTATTCTGAA	AAATAGAGGA	AAGAATACTT	CCAACTCAT	-10014
TCTACATGGC	CAGTATTACC	CTGATTCCAA	AACCAGACAA	AAACACATCA	AAAACAAACA	AACAAAAAAA	-9944
CAGAAAGAAA	GAAAACTACA	GGCCAATATC	CCTGATGAAT	ACTGATACAA	AAATCCTCAA	CAAAACACTA	-9874
GCAAAACCAA	TTAAACAACA	CCTTCGAAAG	ATCATTCATT	GTGATCAAGT	GGGATTTTAT	CCAGGGATGG	-9804
AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAAATGAAGT	ACAAAACTA	-9734
TATGATTATT	TCACTTTATG	CAGAAAAAGC	ATTTGATAAA	ATTCTGCACC	CTTCATGATA	AAAACCCTCA	-9664
AAAAACCAGG	TATACAAGAA	ACATACAGGC	CAGGCACAGT	GGTCCACACC	TGCGATCCCA	GCACTCTGGG	-9594
AGGCCAAGGT	GGGATGATTG	CTTGGGCCCC	GGAGTTTGAG	ACTAGCCTGG	GCAACAAAAT	GAGACCTGGT	-9524
CTACAAAAAA	CTTTTAAAAA	AAATTAGCCA	GGCATGTGAG	CATATGCCTG	TAGTCCCAGC	TAGTCTGGAG	-9454
GCTGAGGTGG	GAGAATCACT	TAAGCCTAGG	AGGTGCAGGC	TGCAGTGAGC	CATGAACATG	TCAGTGTACT	-9384
CCAGCCTAGA	CAACAGAACA	AGACCCCACT	GAATAAGAAG	AAGGAGAAGG	AGAAGGGAGA	AGGGAGGGAG	-9314
AAGGGAGGAG	GAGGAGAAGG	AGGAGGTGGA	GGAGAAGTGG	AAGGGGAAGG	GGAAGGGAAA	GAGGAAGAAG	-9244
AAGAAACATA	TTTCAACATA	ATAAAGCCCC	TATATGACAG	ACCGAGGTAG	TATTATGAGG	AAAACTGAA	-9174
AGCCTTTCCCT	CTAAGATCTG	GAAAAAGACA	AGGGCCCACT	TTCAACCACTG	TGATTCAACA	TAGTACTAGA	-9104
AGTCTTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAAA	GGCATCCAAA	CTGGAAAGGA	AGAAGTCAAA	-9034
TTATCCTGTT	TGCAGATGAT	ATGATCTTAT	ATCTGGAAGA	GACTTAAGAC	ACCACTAAAA	AACTATTAGA	-8964
GCTGAAATTT	GGTACAGCAG	GATACAAAAT	CAATGTACAA	AAATCAGTAG	TATTTCTATA	TTCCAACAGC	-8894
AAACAATCTG	AAAAAGAAAC	CAAAAAAGCA	GCTACAAATA	AAATTAACA	GCTAGGAATT	AACCAAGAA	-8824
GTGAAAGATC	TCTACAATGA	AACTATAAAA	ATGTTGATAA	AAGAAATTGA	AGAGGGCACA	AAAAAAGAAA	-8754
AGATATTCCA	TGTTCATAGA	TTGGAAGAAT	AAATACTGTT	AAAATGTCCA	TACTACCCAA	AGCAATTAC	-8684
AAATTCAATG	CAATCCCTAT	TAAATACTA	ATGACGTTCT	TCACAGAAAT	AGAAGAAACA	ATTCTAAGAT	-8614
TTGTACAGAA	CCACAAAAGA	CCGAGAATAG	CCTGACCAAT	TCTGACCAAA	AAGAACAACA	CTGGAAGCAT	-8544
CACATTACCT	GACTTCAAT	TATACTACAA	AGCTATAGTA	ACCCAAACTA	CATGGTACTG	GCATAAAAC	-8474
AGATGAGACA	TGGACCAGAG	GAACAGAATA	GAGAATCCAG	AAACAAATCC	ATGCATCTAC	AGTGAACCTA	-8404
TTTTTGACAA	AGGTGCCAAG	AACATACTTT	GGGGAAGAGA	TAATCTCTTC	AATAAATGGT	GCTGGAGGAA	-8334
CTGGATATCC	ATATGCAAAA	TAACATACT	AGAATCTGT	CTCTCACCAT	ATACAAAAGC	AAATCAAAAT	-8264
GGATGAAAGG	CTTAAATCTA	AAACCTCAAA	CTTTGCAACT	ACTAAAAGAA	AACACCGGAG	AAACTCTCCA	-8194
GGACATTGGA	TGTGGCAAAG	ACTTCTTGAG	TAATCTCCCTG	CAGGCACAGG	CAACCAAAGG	AAAAACAGAC	-8124
AAATGGGATC	ATATCAAGTT	AAAAAGCTTC	TGCCCAAGCA	AGGAAACAAT	CAACAAAGAG	AAGAGACAAC	-8054
CCACAGAATG	GGAGAATATA	TTTGCAAACT	ATTCTATCTAA	CAAGGAATTA	ATAACCAGTA	TATATAAGGA	-7984
GCTCAAACTA	CTCTATAAGA	AAAACACCTA	ATAAGCTGAT	TTTCAAAAAT	AAGCAAAAGA	TCTGGGTAGA	-7914
CATTTCTCAA	AATAAGTCAT	ACAAATGGCA	AACAGGCATC	TGAAAATGTG	CTCAACACCA	CTGATCATCA	-7844
GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCCAGTTAAA	ATGGCTTTTA	TTCAAAAGAC	-7774
AGGCAATAAC	AAATGCCAGT	GAGGATGTGG	ATAAAAGGAA	ACCTTGGAC	ACTGTTGGTG	GGAAATGAAA	-7704
TTGCTACCAC	TATGGAGAAC	AGTTTGAAAG	TTCTCTAAAA	AACCTAAAAAT	TAAGCTACCA	TACAGCAATC	-7634
CCATTGCTAG	GTATATACTC	CAAAAAAGGG	AATCAGTGTA	TCAACAAGCT	ATCTCCACTC	CCACATTTAC	-7564
TGCAGCACTG	TTCATAGCAG	CCAAGGTTTG	GAAGCAACCT	CAGTGCCAT	CAACAGACGA	ATGGAAAAAG	-7494
AAAATGTGGT	GCACATACAC	AATGGAGTAC	TACGCAGCCA	TAAAAAAGAA	TGAGATCCTG	TCAGTTGCAA	-7424
CAGCATGGGG	GGCACTGGTC	AGTATGTTAA	GTGAAATAAG	CCAGGCACAG	AAAGACAAAC	TTTTCATGTT	-7354
CTCCCTTACT	TGTGGGAGCA	AAAATTAAAA	CAATTGACAT	AGAAATAGAG	GAGAATGGTG	GTTCTAGAGG	-7284
GGTGGGGGAC	AGGGTGACTA	GAGTCAACAA	TAAATTTATTG	TATGTTTTAA	AATAACTAAA	AGAGTATAAT	-7214
TGGGTTGTTT	GTAACACAAA	GAAAGGATAA	ATGCTTGAAG	GTGACAGATA	CCCCATTTAC	CCTGATGTGA	-7144
TTATTACACA	TTGTATGCCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCTTA	CTATATTAAA	-7074
AATTAATAAT	TTAATGGCCA	GGCACGGTGG	CTCATGTCCG	TAATCCCAGC	ACTTTGGGAG	GGCGAGGCGG	-7004
GTGGATCACC	TGAGGTCAGG	AGTTTGAAAC	CAGTCTGGCC	ACCATGATGA	AACCCTGTCT	CTACTAAAGA	-6934
TACAAAAAAT	AGCCAGGCGT	GGTGGCACAT	ACCTGTAGTC	CCAATCTACT	AGGAGGCTGA	GACAGGAGAA	-6864
TTGCTTGAAC	CTGGGAGGCG	GAGGTTGCAG	TGAGCCGAGA	TCTGCCCCT	GCACTGCAGC	CTGGGTGACA	-6794
GAGCAAGACT	CCATCTCAAA	ACAAAAACAA	AAAAAAGAA	ATTAAAAATG	TAATTTTTAT	GTACCGTATA	-6724
AATATATACT	CTACTATATT	AGAAGTTAAA	AATTAACA	ATTATAAAG	GTAATTAACC	ACTTAATCTA	-6654
AAATAAGAAC	AATGTATGTG	GGGTTTCTAG	CTTCTGAAGA	AGTAAAGTT	ATGCCACGA	TGGCAGAAAT	-6584

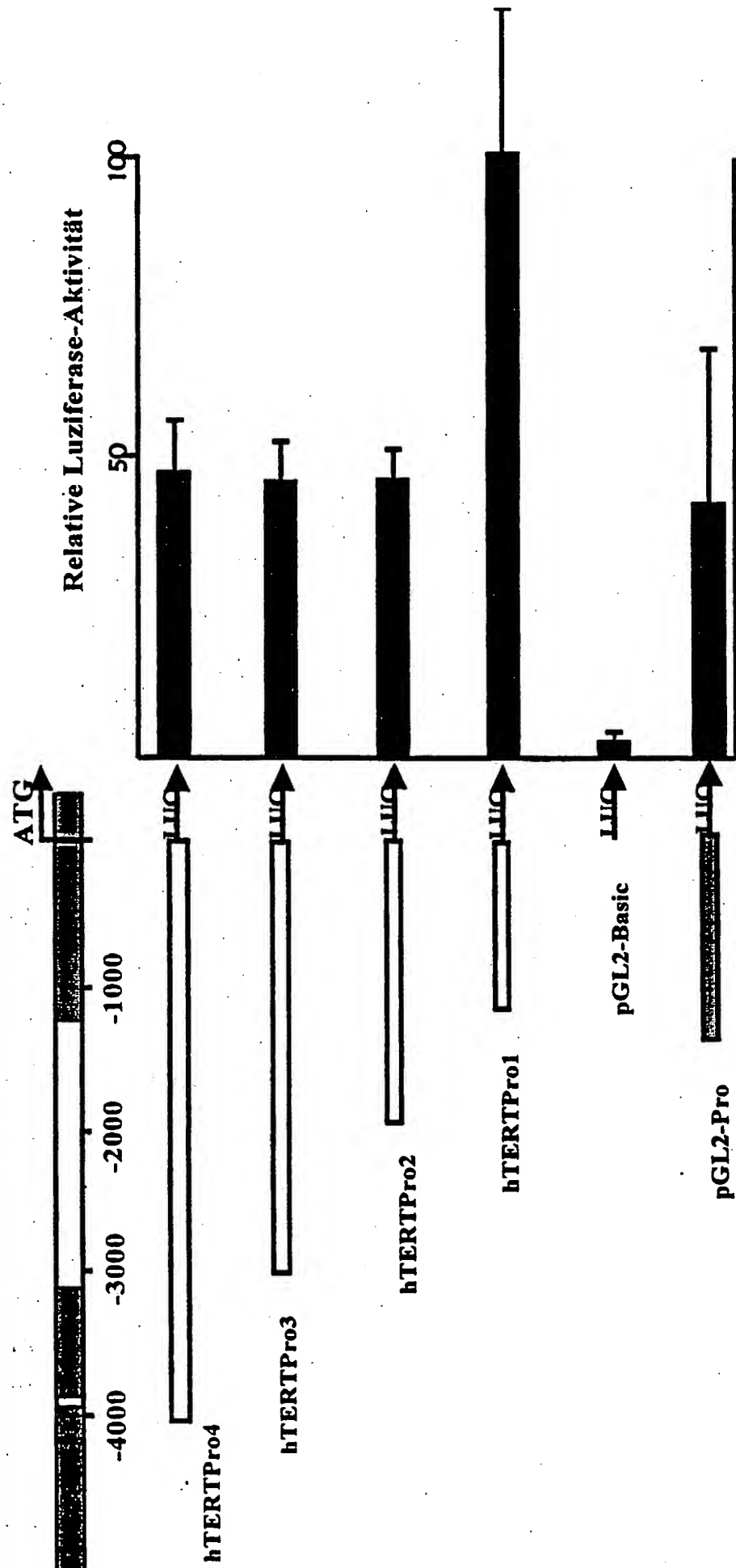
Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCCTC TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGAAA -6024
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAAACAGAA AGTCATGGAA -5814
 GCACCTTCT CAAGGGA AAAACAGGAA CTTGAGTTAG GTGCCTTCTT TAAAACAGAA AGTCATGGAA -5814
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCCGA GCGGCTGCTG AGGACCTCT -5744
 TGCAAGAGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTC AACCTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTACAGG CTGCAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCG CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAG GTACACGAGG -5324
 AGAGGCGTGG CGGCGAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184
 GTGACTCAGG ACCCATACC GGCTTCCTGG GCCCACCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAAC GAACATGACC CTGCGCTGCC TGCTTCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTCGA -5044
 GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCACC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA AGGCCATGA CAAAACCTAG TACAAACACC -4904
 ACTCTTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTACGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTC TCTAGACTAG TAGACCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT CTCAGCAGC -4624
 TTCTTAAACC CTGGGTGGGC CGTGTTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGCTCAGCG -4554
 ACGTAGCTCG CACGGTTCTT CCTCACATGG GGTGTCTGTC TCCTTCCCA AACTCACAT GCGTTGAAGG -4484
 GAGGAGATT TCGCCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGAGGTT -4414
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG CCGGGGCTG -4344
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCC CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GCGGCTCTTG GGAATGCAA CATTGGGTG -4204
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCCGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGACACC CTGGACATT GCCCCACAGC CCTGGGAATT -3994

 c-Myc
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAAATTTT ACCCCATGGC -3714
 AGGGGAGTGG TTAGGGGGGT TAAGGACCGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644
 AAGCCAGTTT CTTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCCC -3504
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACC CGCCG GCGCCAGGGC CTTTGCAGGT GTGATCTCCG -3434
 TGAGGACCTT GAGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCGAAAA GTAATCCAGG GGTTCCTGGGA -3364
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAG GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCTCC -3154
 GTGCCATAGG AGGGCACTCG CGTGCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
 GTGCTCCCG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGCCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTTCAGGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCCTATTTA ACCATTTTAA AACTTCCTG GGCTCAAGTC ACACCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTTCCCCT TTAATCAGGA GTTACCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
 CGTCTCTGA CATATTCACA GTTTCGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCG AGGAGTTCCT CTCCTCCTG -2244
 TGGAGGAAGG AATGATACTT TGTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGTTTGG -2174
 TTTGTTTTGT TTTGAGAGGC GGTTCCTC TTTGTGCTCA GGCTGGAGGG AGTGCAATGG CGCATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCG CTCCATTG GCTGGGATTA -2034
 CAGGACCCG CCACCATGCC CAGCTAATT TTTGTATTT TAGTAGAGAC GGGGGTGGG GGGGTTCCAC -1964

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824
CAAT-Box
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTATTGT TGTTAGAACA -1754
CTCTTGATGT TTTACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCAATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTTCT TCTCTCCCT CTTTTAAAT TGTGTTTTCT -1544
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTTGT TGGAACAAAT TTTCCAACC -1474
Sp1
GCCCTTTGC CCTAGTGGCA GAGACAATC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404
GGATTTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
GCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTCCTCC GGCAGTTTCT -1264
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCAAGTGT TTGCCGACCT CAGCTACAGC -1194
ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CTTAGATCC AAAGTTGAGC AACCCGGAGT -1124
CTGGATTCTT GGGAAAGTCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG, GGACCAGTGG -1054
CGTGTTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
GCCTGGACCC CGAGGCTGCC CTCCACCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914
ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844
CCAC-Box
GCGCCTGGCT CCATTCCCA CCGTTTCTCG ACGGGACCGC CCCGGTGGT GATTAACAGA TTTGGGGTGG -774
Sp1
TTTGCTCATG GTGGGGACCC CTCGCCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG TGTCAGGAG -704
CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCCTCGT GCCCGTCCAG GGAGCAATGC -634
AP-2
GTCCTCGGGT TCGTCCCCAG CCGCGTCTAC GCGCCTCCGT CCTCCCCTTC ACGTCCGGCA TTCGTGGTGC -564
CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
GGTCGCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCTCCCT CGGGTTACCC CACAGCCTAG -424
GCCGATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA GCGGGGCGCG -354
Sp1
GCGCGGGAAG CGCGGCCAG ACCCCCGGGT CGGCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284
CCAGTGGAAT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CACGTGCG GAGGGACTGG GGACCCGGGC -214
c-Myc
ACCGTCTCTG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACCC GCCCGTCCC GACCCCTCCC -144
Sp1
GGGTCCCCG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTT TCCGCGGCC CGCCCTCTCC -74
c-Myc
TCCGGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCC ACGTGGGAAG CCCTGGCCCC GGCCACCCCC -4
GCGATG

Fig.: 11



SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>

<141>

15 <160> 20

<170> PatentIn Vers. 2.0

<210> 1

20 <211> 5126

<212> DNA

<213> Homo sapiens

<400> 1

25 gagctctgaa ccgtggaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaatgaa gtgggtgtgca ggaaatggcc atgtaaatta cactgactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaacacacc actcttttac taggcccaca 240
gagcacgggc cacaccctg atatattaag agtccaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggg tgctgcttcc cgaggggcgc atctgcccgc 420
gagactcagc ctgggggtgcc acactgaggc cagccctgtc tccacaccct ccgcctccag 480
gcctcagctt ctccagcagc ttccctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cacggttctt cctcacatgg 600
ggtgtctgtc tccttcccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720
ggctgcacgc tgacctccat ttccaggcgc tccccgtctc ctgtcatctg ccggggcctg 780
ccggtgtgtt cttctgtttc tgtgtctcct tccacgtcca gctgcgtgtg tctctgcccg 840
ctaggggtctc ggggttttta taggcatagg acgggggctt ggtggggcag ggcgctcttg 900
ggaaatgcaa catttggttg tgaaagttag agtgctgtc ctcacctagg tccacgggca 960
40 caggcctggg gatggagccc ccgccaggga cccgcccttc tctgcccagc actttcctgc 1020
ccccctccct ctggaacaca gagtggcagt ttccacaagc actaagcatc ctcttcccaa 1080
aagaccacgc attggcaccc ctggacattt gccccacagc cctgggaatt cacgtgacta 1140
cgacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatatgta 1200
caaaagcagg aaatccctgc taaaatgtcc tttaacaaac tggttaaaca aacgggtcca 1260
45 tccgcacggg ggacagttcc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtggt ttaggggggt taaggacggt 1440
gggggaggca gctgggggct actgcacgca ctttttacta aagccagttt cctgggtctg 1500
atggtatttg ctctgttatg ggagactaac cataggggag tggggatggg ggaaccggga 1560
50 ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataatgctc agagatgcc 1620
acgtcctgat tccccaaac ctgtggacag aaccgcggcg gccccagggc ctttgaggt 1680
gtgatctccg tgaggacctt gaggtctggg atccttcggg actacctgca ggcccgaaaa 1740
gtaatccagg ggttctggga agaggcgggc aggggggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggaggaggag cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcagggaag 1920
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactgc cgctgcctt ctagcatgaa gtgtgtggg atttgagaa gcaacaggaa 2040
acctatgcac tgtgaactta ggattatttc aaaaacaaagg ttacagaaa catccaagga 2100
cagggctgaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
60 ttattttatt tacttacttt ctgagacaga gttatgctct tgttgcccag gctggagtgc 2220
agcggcatga tcttggtcct ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccggctt aattttgtat 2340
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaaatc ctgacctcag 2400
gtgatccgcc cacctcagcc tcccaaagtg ctgggattac aggcattgag cactgcacct 2460

5 ggccctattta accatttttaa aacttccctg ggetcaagtc acacccactg gtaaggagtt 2520
 catggagttc aatttcccct ttactcagga gttaccctcc tttgatattt tctgtaattc 2580
 ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
 tcccattggga cccactgcag gggcagctgg gaggctgcag gcttcaggtc ccagtggggt 2700
 tgccatctgc cagtagaaac ctgatgtaga atcaggggcg aagtgtggac actgtcctga 2760
 atctcaatgt ctacgtgtgt gctgaacat gtagaaatta aagtcctacc ctccactct 2820
 actgggattg agcccttcc ctatccccc ccaggggcag aggagttcct ctactcctg 2880
 tggaggaagg aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940
 10 tgttggtttg tttgttttgt tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000
 agtgcaatgg cgcgatcttg gcttactgca gcctctgcct ccaggttca agtgattctc 3060
 ctgcttccgc ctcccatttg gctgggatta caggcaccgc ccaccatgcc cagctaattt 3120
 tttgtatttt tagtagagac ggggttgagg ggggttcacc atgttggtcca ggctgggtctc 3180
 gaactcaatg cctcagatga tccacctgcc tctgcctcct aaagtgcctg gattacaggt 3240
 15 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
 gaagctcacc ccactcaagt gttgtggtgt ttttaagcaa tgatagaatt tttttattgt 3360
 tgttagaaca ctcttgatgt tttacactgt gatgactaag acatcatcag cttttcaaag 3420
 acacactaac tgcacccata ataactgggt gtcttctggg tatcagcaat ctctattgaa 3480
 tgccgggagg cgtttcctcg ccactgcacat ggtgttaatt actccagcat aatcttctgc 3540
 20 tcccatttct tctcttccct cttttaaaaat tgtgttttct atgttggtct ctctgcagag 3600
 aaccagtgtg agctacaact taacttttgt tggaaacaaat tttccaaacc gcccttttgc 3660
 cctagtggca gagacaattc acaaacacag ccctttaaaa aggccttaggg atcactaagg 3720
 ggattttctag aagagcgacc tgtaatccta agtatttaca agacgaggct aacctccagc 3780
 gagcgtgaca gccaggggag ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
 25 aatttctctc ggagtttct gaaagtagga aaggttacat ttaaggttgc gtttggttagc 3900
 atttctagtg ttgcccagct cagctacagc atccctgcaa ggcctcggga gaccagaagc 3960
 tttctcgccc ccttagatcc aaacttgagc aaccggaggt ctggattcct gggaagtcct 4020
 cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
 tctactgtcg ggctggaagt cgggcctcct agctctgcag tccgaggctt ggagccaggc 4140
 30 gcctggaccc cgaggctgcc ctccaccctg tgcggcgagg atgtgaccag atgttgacct 4200
 catctgccag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
 ccggtgcgcg gccagcagga gcgcctggct ccatttccca ccctttctcg acgggaccgc 4320
 ccggtgggtg gattaacaga tttgggtggg tttgctcatg gtggggaccc ctgcgcctc 4380
 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtcgc ggggaagtgt 4440
 35 tgcagggagg cactccggga ggtcccgctg gcccgtccag ggagcaatgc gtccctcgggt 4500
 tcgtccccag ccgctcttac gcgcctcctg cctcccttcc acgtccggca ttcgtggtgc 4560
 ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
 gcggccaaag ggtgcgcgca cgcacctgtt cccagggcct ccacatcatg gccctccct 4680
 cgggttacc cagagcctag gccgattcga cctctctcct ctggggccct cgctggcgtc 4740
 40 cctgcacctt gggagcgcga gcgcgcgcg ggcggggaag cgcgggccag acccccggt 4800
 ccgcccggag cagctgcgct gtcggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggacccgggc acccgtctcg 4920
 ccccttacc ttccagctcc gcctcctcct cgcggacccc gcccctccc gaccctccc 4980
 ggggtccccg cccagcccc tccgggcccct cccagcccc ccccttctct tccgcgcccc 5040
 45 cgccctctcc tcgcggcgcg agtttcaggc agcgtgcgt cctgctgcgc acgtgggaag 5100
 ccctggcccc ggccaccccc gcgatg 5126

<210> 2

<211> 4042

<212> DNA

50 <213> Homo sapiens

<400> 2

55 gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
 cgatgccgcg cgtccccgcg tgcgagccg tgcgctccct gctgcgcagc cactaccgcg 120
 aggtgtgtcc gctggccacg ttcgtgcggc gcctggggcc ccagggtgtg cggctggtgc 180
 agcgcgggga cccggcggtt ttccgcgcgc tgggtggcca gtgcctggtg tgcgtgccct 240
 gggacgcacg gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300
 tgggtggccc agtgcgtgag aggtgtgctg agcgcggcgc gaagaacgtg ctggccttcg 360
 60 gcttcgcgct gctggacggg gcccgcgggg gccccccga ggccttcacc accagcgtgc 420
 gcagctacct gcccaacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
 tgctgcgcgg cgtgggagac gacgtgctgg ttcacctgct ggacgcgtgc gcgctctttg 540
 tgctgggtgg tcccagctgc gcctaccagg tgtgcggggc gccgctgtac cagctcggcg 600
 ctgccactca ggcccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660
 aacgggcctg gaaccatagc gtcagggagg ccggggctcc cctgggcctg ccagcccccg 720
 65 gtgcgaggag gcgcgggggc agtgcacgcc gaagtctgcc gttgccaag aggccaggc 780

3 / 18

5 gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gccacccgg 840
 gcaggacgcg tggaccgagt gaccgtggtt tctgtgtggt gtcacctgcc agaccgcgg 900
 aagaagccac ctctttggag ggtgcgctct ctggcagcg ccactccac ccatccgtgg 960
 gccgccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020
 10 ettggtcccc ggtgtacgcc gagaccaagc acttcctcta ctctcaggg gacaaggagc 1080
 agctgcggcc ctcttctcta ctacgtctc tgaggcccg cctgactggc gctcggaggc 1140
 tcgtggagac catctttctg ggttccaggg cctggatgcc agggactccc cgcaggttgc 1200
 cccgctgccc ccagcgctac tggcaaatgc ggccctggt tctggagctg cttgggaacc 1260
 acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctgcca gctgcggtca 1320
 cccagcagc cggtgtctgt gcccgggaga agccccagg ctctgtggcg gccccgagg 1380
 aggaggacac agacccccgt cgctgtgctc agctgctccg ccagcacagc agccccggc 1440
 aggtgtacgg ctctgtgcgg gcctgcctgc gccggctggt gccccaggc ctctggggct 1500
 ccaggcaca cgaacgcgc ttctcagga acaccaagaa gttcatctcc ctggggaagc 1560
 15 atgccagct ctgcgtcag gagctgacgt ggaagatgag cgtgcgggac tgcgcttggc 1620
 tgccgaggag cccaggggtt ggtgtgttcc cggccgcaga gcacctctg cgtgaggaga 1680
 tcttgcccaa gttctgcac tggctgatga gtgtgtacgt cgtcgagctg ctgaggtctt 1740
 tcttttatgt caggagacc acgtttcaaa agaaccaggct cttttctac cgggaagagt 1800
 tctggagcaa gttgcaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860
 20 agctgtcggg agcagaggtc aggcagcatc gggaagccag gccccctg ctgacgtcca 1920
 gactccgctt catccccaa cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg aggggtgaag 2040
 cactgttcag cgtgctcaac tacgagcggg cgccgcccgc cggcctcctg ggcgcctctg 2100
 tgctgggcct ggacgatata cacaggcctt ggcgacactt cgtgctgctg gtgcccggcc 2160
 25 aggaccgcgc gcctgagctg tactttgtca aggtggatgt gacgggcgcg tacgacacca 2220
 tccccagga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgcgtcg gtatgccgtg gtccagaagg ccgccatgg gcacgtccgc aaggccttca 2340
 agagccacgt ctctaccttg acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
 tgcaggagac cagcccgtcg agggatgccg tcgtcatcga gcagagctcc tccctgaatg 2460
 30 agccagcagc tggcctcttc gacgtcttcc tacgttctat gtgccaccac gccgtgcgca 2520
 tccggggcaa gctctacgtc cagtgccagg gcatcccgca gggctccatc ctctccacgc 2580
 tgctctgcag cctgtgttac ggcgacatgg agaacaagct gtttcggggg attcggcggg 2640
 acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760
 35 tgcggaagac agtgggtgaa ttccctgtag aagacgaggc cctgggtggc acggcttttg 2820
 ttcagatgcc ggcccacggc ctattcccct ggtgcggcct gctgctggat acccgaccc 2880
 tgaaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcact 2940
 tcaaccgcgg ctccaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgccgc 3000
 tgaagtgtca cagcctgttt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060
 40 acatctacaa gatectctg ctgcaggcgt acaggtttca cgcattgtgt ctgcagctcc 3120
 catttcatca gcaagtttgg aagaacccca catttttctt gcgctcatc tctgacacgg 3180
 cctccctctg ctactccatc ctgaaagcca agaaccgagg gatgtcgtg ggggccaagg 3240
 tgcgcggcgg ccctctgccc tccgaggccg tgcagtggct gtgccaccaa gcattcctgc 3300
 gcaagctgac tcgacaccgt gtcacctcct tgccactcct ggggtcactc aggacagccc 3360
 45 agacgcagct gagtcggaag ctcccgggga cgacgtgac tgccctggag gccgcagcca 3420
 acccggcact gccctcagac ttcaagacca tctggactg atggccacc gccacagcc 3480
 aggcgagag cagacaccag cagccctgtc acgcggggt ctacgtccca gggaggagg 3540
 ggcggccac acccaggccc gcaccgctgg gactctgagg cctgagttag tgtttggcg 3600
 aggcctgcat gtccggtga aggtgagtg tccggtgag gcctgagcga gtgtccagcc 3660
 50 aagggtgtag tgtccagcac acctgcctc ttacttccc cacaggctg cgctcggctc 3720
 caccacggg ccagcttttc ctaccaggga gcccggttc cactccccac ataggaatag 3780
 tccatcccca gattcgccat tgttaccctc tcgcccctgc ctcttttgc ttccaccccc 3840
 accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gactgaccaa 3900
 aggtgtgccc tgtacacagg cgaggacct gcacctggat gggggctcct gtgggtcaaa 3960
 55 ttgggggggag gtgctgtggg agtaaaatac tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3

<211> 11276

<212> DNA

<213> Homo sapiens

<400> 3

60 acttgagccc aagagttcaa ggctacgggt agccatgatt gcaacaccac acgccagcct 60
 tgggtgacaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120
 65 tcttctcttg ccacagtggg acaaaaccag aaatcaacaa caaggagaa tttgaaaact 180

5 atacaaacac atgaaaatta aacaatatatc ttctgaatga ccagtgaagtc aatgaagaaa 240
 ttaaaaaagga aattgaaaaa tttattttaag caaatgataa cggaaacata acctctcaaa 300
 acccacggta tacagcaaaa gcagtgtctaa gaaggaagtt tatagtctata agcagctaca 360
 tcaaaaaagt agaaaagcca ggcgcagtgg ctcatgcctg taatcccagc actttgggag 420
 5 gccaaggcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480
 aaccttgtcg ctactaaaaa taaaaaatta gctgggcatg gtggcacatg cctgtaatcc 540
 cagctactcg ggaggctgag gcaggataac cgcttgaacc caggaggtgg aggttgcggt 600
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660
 10 aaaaaaaaaa aagttagaaa acttaaaaat acaacctaata gatgcacctt aaagaactag 720
 aaaagcaaga gcaaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780
 cagaaataaa tgaaactgaa agataacaat acaaaagatc aacaaaatta aaagtgtggt 840
 ttttgaaaag ataaacaaaa ttgacaaacc tttgcccaga ctaagaaaaa aggaaagaa 900
 acctaataaa ataaagtcag agatgaaaaa agagacatta caactgatac cacagaaatt 960
 15 caaaggatca cttagaggcta ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020
 aaaatagata aattcctaga tgcatacaac ctaccaagat tgaacctaga agaaatccaa 1080
 agcccaaca gaccaataac aataatggga ttaaagccat aataaaaagt ctctagcaa 1140
 agagaagccc aggacccaat ggcttccttg ctggatttta ccaatcattt aaagaagaa 1200
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaataact ccaaacat 1260
 20 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaacaaca 1320
 aacaaaaaaa cagaaagaaa gaaaactaca ggccaatatc cctgatgaat actgatacaa 1380
 aaatcctcaa caaaacacta gcaaaccaaa ttaaacaca ccttcgaaag atcattcatt 1440
 gtgatcaagt gggatttatt ccaggatggg aaggatggtt caacatatgc aaatcaatca 1500
 atgtgataca tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcaactttatg 1560
 25 cagaaaaagc atttgataaa attctgcacc ctcatgata aaaaacctca aaaaaccagg 1620
 tatacaagaa acatacaggc caggcacagt ggctcacacc tgcgatccca gcactctggg 1680
 agccaaggt gggatgattg cttgggcccc ggagtttgag actagcctgg gcaacaaaat 1740
 gagacctggg ctacaaaaaa cttttttaaa aaattagcca ggcatgatgg catatgcctg 1800
 tagtcccagc tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcgaggc 1860
 30 tgcagtgaag catgaacatg tcaactgtact ccagcctaga caacagaaca agaccccat 1920
 gaataagaag aaggagaagg agaagggaag agggaggag aaggaggag gaggagaag 1980
 aggggtgga ggagaagtgg aaggggaaag ggaagggaag gaggaaagaa aagaaacata 2040
 tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100
 agcctttcct ctaagatctg gaaaatgaca agggccact ttcaccactg tgattcaaca 2160
 35 tagtactaga agtcctagct agagcaatca gataagagaa agaaataaaa ggcacccaaa 2220
 ctggaaagga agaagtcaaa ttatcctgtt tgcagatgat atgatcttat atctggaaa 2280
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaa 2340
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaac 2400
 40 caaaaaagca gctacaaata aaattaaaca gctaggaatt aaccaagaa gtgaaagatc 2460
 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaaagaaa 2520
 agatattcca tgttcataga ttggaagaat aaatactgtt aaaatgtcca tactaccaa 2580
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640
 agaagaaaca atttcaagat ttgtacagaa ccacaaaaga cccagaatag ccaaagctat 2700
 cctgaccaa aagaacaaaa ctggaagcat cacattacct gacttcaat tatactaca 2760
 45 agctatagta acccaacta catggtactg gcataaaaaac agatgagaca tggaccagag 2820
 gaacagaata gagaatccag aaacaatctc atgcatctac agtgaaacta tttttgaca 2880
 aggtgccaag aacatacttt ggggaaaaa taatctcttc aataaatggt gctggaggaa 2940
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcaccat atacaaaagc 3000
 aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcact actaaaagaa 3060
 50 aacaccggag aaactctcca ggacattgga gtgggcaag acttcttgag taattccctg 3120
 caggcacagg caaccaagc aaaaacagac aaatgggatc atatcaagtt aaaaagcttc 3180
 tggccagcaa aggaaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240
 tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300
 ctctataaga aaaacaccta ataagctgat tttcaaaaat aagcaaaaga tctgggtaga 3360
 55 catttctcaa aataagtcac acaaatggca aacaggcatc tgaaaatgtg ctcaacacca 3420
 ctgatcatca gagaaatgca aatcaaaact actatgagag atcatctcat cccagttaaa 3480
 atggctttta ttcaaaagac aggaataaac aaatgccagt gaggatgtgg ataaaaggaa 3540
 acccttggac actgttgggt ggaatgaaa ttgctaccac tatggagaac agtttgaaag 3600
 60 ttcctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatctc 3660
 caaaaaaggg aatcagtgtg tcaacaagct atctccactc ccacatttac tgcagcactg 3720
 ttcatagcag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780
 aaaaatgtgg gcacatacac aatggagtac tacgcagcca taaaaaagaa tgagatcctg 3840
 tcagttgcaa cagcatgggg ggcaactggt agtatgttaa gtgaaataag ccaggcacag 3900
 aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaaattaaaa caattgacat 3960
 65 agaaatagag gagaatgggt gttctagagg ggtgggggac aggggtgacta gagtcaacaa 4020
 taatttattg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

5 gaaaggataa atgcttgaag gtgacagata cccattttac cctgatgtga ttattacaca 4140
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
 aattaaaatt ttaattggcca ggcacgggtg ctcacgtccg taatcccagc accttggggag 4260
 gccgagggcg gtggatcacc tgaggtcagg agtttgaac cagtctggcc accatgatga 4320
 aacctgtct ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgcag 4440
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
 aaaaaacaa aaaaaagaag attaaaattg taatttttat gtaccgtata aatatatact 4560
 ctactatatt agaagttaaa aattaaaaca attataaaag gtaattaacc acttaattcta 4620
 10 aaataagaac aatgtatgtg ggggtttctag cttctgaaga agtaaaagt atggccacga 4680
 tggcagaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctg 4740
 taagtgaact aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800
 cctaaaacaa ctgctaataa tgggtgaaagg taatctctat taattaccaa taattacaga 4860
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtctc tcattcacgg 4920
 15 tgcctttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980
 atctgtatga atctgaaac gaaaaatggt ggtgatttcc tccagaagaa ttagagtacc 5040
 tggcaggaag caggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
 aagacccaag tgcaaggcag aggcctgatg acccgaggac aggaaaagctc ggatgggaag 5160
 gggcgatgag aagcctgcct cgttgggtgag cagcgcatga agtgccctta tttacgcttt 5220
 20 gcaaagattg ctctggatc catctggaaa aggcggccag cggaatgca aggagtcaga 5280
 agcctcctgc tcaaaaccag gccagcagct atggcgccca cccggcgctg tgcagagggg 5340
 agaggagtca aggcacctcg aagtatggct taaatctttt tttcacctga agcagtgaac 5400
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatggaa 5460
 gcacctctt caagggaaaa ccagacgccc gctctgcggt catctacctc tttcctctct 5520
 25 cctctctctg cctcgcggt tctgtatcgg gacagagtga ccccgctgga gcttctccga 5580
 gcccgctgctg aggacctctc tgcaaaagggc tccacagacc cccgccttg agagaggagt 5640
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760
 tcttaattatt ttcttaaaat tcatcaaaata acattcagga ctgcagaaat ccaaaaggcgt 5820
 30 aaaaacaggaa ctgagctatg tttgccaagg tccaaggact taataacat gttcagaggg 5880
 attttctgct ctaagtactt tttattggtt ttcataaggt ggcttagggg gcaagggaag 5940
 gtacacgagg agaggcctgg ggcgcagggc tatgagcacg gcagggccac cggggagaga 6000
 gtccccggcc tgggaggctg acagcaggac cactgaccgt cctccctggg agctgccaca 6060
 35 ttgggcaacg cgaaggcggc cacgtgcgt gtgactcagg accccatacc ggcttctg 6120
 gccacccac actaaccag gaagtacagg agctctgaac ccgtggaac gaacatgacc 6180
 cttgctgcc tgcttccctg ggtgggtcaa gggtaatgaa gtggtgtgca ggaatggcc 6240
 atgtaatta cagactctg ctgatggga cggttcctc catcattatt catcttacc 6300
 ccaaggact gaatgattcc agcaactct tccggtgtga caagccatga caaaactcag 6360
 40 taaaaacacc actcttttac taggcccaca gagcacggsc cacaccctg atatattaag 6420
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
 agtctgttcc tctagactag tagaccctgg caggcactcc cccagattct agggcctggt 6540
 tgctgcttcc cgagggcgcc atctgccctg gagactcagc ctgggtgacc aactgaggc 6600
 cagecctgtc tccacacctc ccgctccag gcctcagct ctccagcagc ttcctaaacc 6660
 45 ctgggtgggc cgtgttccag cgtactgtc tcactgttc cactgtgtc tgtctcagc 6720
 acgtagctcg cacggttct cctcacatg ggtgtgtc tcttcccca acactcacat 6780
 gcgttgaagg gaggagattc tgcgcctccc agactggctc ctctgagcct gaacctggct 6840
 cgtggccccc gatgcagggt cctggcgctc ggtgcacgc tgacctccat tccaggcgc 6900
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgttcc tgtgtcctt 6960
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggtctc ggggttttta taggcatagg 7020
 acggggcggt ggtgggccag ggcgctcttg ggaaatgcaa catttgggtg tgaagtagg 7080
 agtgcctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
 ccgcctctc tctgcccagc actttcctgc cccctccct ctggaacaca gaggggcagt 7200
 tccacaagc actaagcatc ctcttcccaa aagaccagc attggcacc ctggacattt 7260
 55 gccccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
 ccgacccccg ctgttttatt ttaatagcta caaagcagg aaatccctgc taaaatgtcc 7380
 tttaacaac tggttaaaac aacgggtcca tccgcacggt ggacagtcc tcacagtga 7440
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gactcaaac 7500
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc acccatggc 7560
 60 aggggagtg ttaggggggt taaggacggt gggggcgga gctgggggt actgcacgca 7620
 ccttttacta aagccagttt cctggtctg atggtattg ctgagttatg ggagactaac 7680
 cataggggag tggggatggg ggaacccgga ggctgtgcca tctttgccat gcccgagtgt 7740
 cctgggcagg ataagtctct agagatgccc acgtcctgat tccccaaac ctgtggacag 7800
 aacccgcccg gccccagggc ctttgcaggt gtgatctccg tgaggacctt gaggctctgg 7860
 65 atccttcggg actacctgca ggcccgaaaa gtaatccagg ggttctggga agaggcgggc 7920
 aggagggta gaggggggca gcctcaggac gatggaggca gtcagctgga ggctgaaaag 7980

6 / 18

5 ggagggaggg cctcgagccc aggcctgcaa ggcctccag aagctggaaa aagcggggaa 8040
 gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100
 atcgtggacc tccggcctcc gtgcatagg agggcactcg cgctgccctt ctagcatgaa 8160
 gtgtgtgggg atttgcaaaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
 5 aaacaaaagg ttacagaaa catccaagga cagggtgaa gtgcctccgg gcaagggcag 8280
 ggcaggcacg agtgatttta ttagctatt ttattttatt tacttacttt ctgagacaga 8340
 gttatgtctt tgtgcccag gctggagtgc agcggcatga tcttggtca ctgcaacctc 8400
 cgtctcctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
 10 gtgaccacc acaccggct aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
 gtcaagctga tctcaaatc ctgacctcag gtgatccgcc cactcagcc tcccaaagt 8580
 ctgggattac aggcagagc cactgcacct ggctatttta accattttaa aacttccctg 8640
 ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700
 gttaccctcc tttgatattt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760
 15 catattcaca gtttctgtga ccactgttta tcccatggga ccactgcag gggcagctgg 8820
 gaggctgcag gcttcaggtc ccagtggggt tgccatctgc cagtagaaac ctgatgtaga 8880
 atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940
 gtagaataa aagtccatcc ctctactct actgggattg agccccctcc ctatccccc 9000
 ccagggcag agggagttcct ctactcctg tggagggaagg aatgatactt tgtattttt 9060
 20 cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgtttgt tttgagaggc 9120
 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgatcttg gcttactgca 9180
 gcctctgcct ccaggttca agtgattctc ctgcttccgc ctccatttg gctgggatta 9240
 caggcacccg ccaccatgcc cagctaattt tttgtatttt tagtagagac ggggggtggg 9300
 ggggttcacc atgttgcca ggctggctc gaacttctga cctcagatga tccacctgcc 9360
 25 tctgcctcct aaagtgtcg gattacaggt gtgagccacc atgccagct cagaatttac 9420
 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtgggtg 9480
 tttagccaa tgatagaatt ttttatttgt tgttagaaca ctcttgatgt ttactactgt 9540
 gatgactaag acatcatcag cttttcaaag acacactaac tgcaccata atactggggt 9600
 gtcttctggg tatcagcaat ctctattgaa tgccgggagg cgtttcctcg ccatgcacat 9660
 30 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaat 9720
 tgtgttttct atgttggtt ctctgcagta aaccagtgt agctacaact taactttgt 9780
 tggaaacaaat ttccaaacc gccctttgc cctagtggca gagacaattc acaaacacag 9840
 ccctttaaaa aggccttaggg atcactaagg ggatttctag aagagcgacc tghtaatcta 9900
 agtatttaca agacgaggct aacctccagc gagcgtgaca gccaggagg ggtgagaggc 9960
 35 ctgttcaaat gctagctcca taaataaagc aatttccctc ggcagtttct gaaagtagga 10020
 aaggttacat ttaagggtgc gtttgttagc atttcagtgt ttgccgacct cagctacagc 10080
 atccctgcaa ggcctcgga gaccagaag tttctcgccc ccttagatcc aaacttgagc 10140
 aaccggaggt ctggattcct gggaaagtcct cagctgtcct gcggttgtgc cggggcccca 10200
 ggtctggagg ggaccagtgg ccgtgtggct tctactgctg ggctggaagt cgggctcct 10260
 40 agctctgcag tccgaggctt ggagccaggt gcctggacc cgaggctgcc ctccacctg 10320
 tgcgggcggg atgtgaccag atgtggcct catctgccag acagagtgc ggggccagg 10380
 gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440
 ccatttccca ccctttctcg acgggaccgc ccggtgggt gattaacaga tttggggtgg 10500
 tttgctcatg gtggggacc ctgcgcgctc gagaacctgc aaagagaaat gacgggcgt 10560
 45 tgtcaaggag ccaagtgcg ggggaagtgt tgcaggagg cactccggga ggtcccgcgt 10620
 gccgctccag ggagcaatgc gtctcgggt tctccccag ccgctctac gcgcctcgt 10680
 cctcccttcc acgtccggca tctgtggtgc ccggagcccg acgccccgcg tccggacctg 10740
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800
 cccagggcct ccacatcatg gccctccct cgggttacc caccagctag gccgattcga 10860
 50 cctctctccg ctggggccct cgctggcgtc cctgcacct gggagcgcga gcggcgcgcg 10920
 ggcggggaag cgcggcccag acccccgggt ccgcccggag cagctgcgct gtcggggcca 10980
 ggccgggctc ccagtggatt cgcgggcaca gacgccagg accgcgctcc ccacgtggcg 11040
 gagggactgg ggaccgggc acccgctctg ccccttacc ttccagctcc gctcctccg 11100
 cgcggacccc gcccgctccc gaccctccc ggggtcccg cccagcccc tccgggcct 11160
 55 cccagccct ccccttccct tcccgggccc cgcctctcc tcgcggcgcg agtttcaggc 11220
 agcgtgcgt cctgctgcgc acgtgggaag ccctggcccc gggcaccccc gcgatg 11276

 <210> 4
 <211> 104
 <212> DNA
 60 <213> Homo sapiens

 <400> 4
 gtgggcctcc ccggggctcg cgtccggctg ggggtgaggg cggccggggg gaaccagcga 60
 catgcggaga gcagcgagg cgactcaggg cgcttcccc gcag 104

<210> 5
<211> 8616
<212> DNA
<213> Homo sapiens

5

<400> 5

gtgaggaggt ggtggccgtc gaggggccag gccccagagc tgaatgcagt agggggtcag 60
aaaagggggc aggcagagcc ctggtcctcc tgtctccatc gtcacgtggg cacacgtggc 120
10 ttttcgctca ggacgtcgag tggacacggt gatctctgcc tctgctctcc ctctgtcca 180
gtttgcataa acttacgagg ttcaccttca cgttttgatg gacacgcggt ttccaggcgc 240
cgaggccaga gcagtgaaca gaggaggctg ggcgcggcag tggagccggg ttgccggcaa 300
tgggaggaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360
tcctcttcgc aatttcaagg gtgggaatga gagggtggga cgagaacccc ctcttctctg 420
gggtgggagg taagggtttt gcaggtgcac gtggtcagcc aatatgcagg tttgtgttta 480
15 agatttaatt gtgtgttgac ggcaggtgac ggtggctcac gccggtaatc ccagcacttt 540
gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600
ggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcaggtggt tgtgtgctg 660
taatcccgag tacttgagg gctgaggcag gagaatcact tgaaccagg aggcggaggc 720
tgagtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactctgt 780
20 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840
actgttctcc agcacagatc ctggtcccat ctttaggtat gaagagggcc acatgggagc 900
agaggacagc agatggctcc acctgctgag gaaggacag tgtttgtggg tgttcagggg 960
atggtgctgc tgggcccctg cgtgtcccca ccctgtttt ctggatttga tgttgaggaa 1020
cctccgctcc agccccctt tggctccagg tgcctccagg ccctaccgtg gcagctagaa 1080
25 gaagtcacca ttccacccc tccccacaaa ctccaagac atgtaagact tccggccatg 1140
cagacaagga gggtagacct cttggggctc tttttttct ttttttctt ttatgggtggc 1200
aaaagtcata taacatgaga ttggcactcc taacaccgtt ttctgtgtac agtcagaaat 1260
tgctaactcg gcggtgttta cagcaggttg cttgaaatgc tgcgtcttgc gtgactggaa 1320
gtccctaccc atcgaacgyc agctgcctca cacctgctgc ggctcagggt gaccacgccc 1380
30 agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc ctctgttag 1440
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500
aggtcacaat ctgcccctgg cttatgcagg gaggtagggc tgggtcccgg gtgtccctgt 1560
cacgtgcagg gtgagtgagg cgttgcctcc aggtgtccct gtcacgtgta gggtagtgta 1620
ggcgcggccc ccgggtgtcc ctgtcccgtg cagcgtgatt gagggtgtgg ccccggtgtg 1680
35 ccctgtcacg ttagagggtga gtgaggcgcc atcccgggt gtccctgtca cgtgtagggt 1740
gagtgaggcg tgggtcccgg gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800
gggtgtccct gtcacgtgca gggtagtgta ggcgcggctc ccgggtgtcc ctctcagggt 1860
tagggtcagg gaggcgcgcc cccaggggtg cctgtcacg ttaggggtga gtgaggcacc 1920
gtccctgggt gtccctccca ggtatagggt gaggtaggca ctgtcccgg gtgtccctgt 1980
40 cacgtgcagg gtgagtgagg cgcggccccc ggggtgtccct ctccaggtgca gggtagtgta 2040
ggcgtgtcc ctgggtgtcc ctgtctcgtg tagggtgagt gaggctctgt ccccgaggtg 2100
ccttggtggt tgcctcactg agcttgcctc tgaatgtttg ctctttctat agccacagct 2160
gcgcgggttg cccattgcct gggtagatgg tgcaggcgca gtgctggtcc ccaagcctat 2220
cttttctgat gctcggctct tcttgggtcac ctctccgttc cattttgcta cggggacacg 2280
45 ggactgcagg ctctgcctc ccgctgtgca ggcactgcag ccacagcttc aggtccgctt 2340
gcctctgttg ggcctggctt gctcaccacg tgcccggcac atgcatgctg ccaatactcc 2400
tctcccagct tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccacgtg 2460
ttgctggaga catcccagaa aggtttctct gtgcccgtga ggaaagcaag tcacccacagc 2520
cccctcactt gtccctgttt ctcccaagct gccctctgct tgggccccct tgggtgggtg 2580
50 gcaacgcttg tcacctatt ctgggcacct gccgtcatt gcttaggctg ggctctgcct 2640
ccagtcgccc cctcacatgg attgacgtcc agccacagg tggagtgtct ctgtctgtct 2700
cctgctctga gacccacgtg gaggcgccgt gtctccgcca gcctctgca gacttccctc 2760
ttgggtctta gttttgaatt tcaactgatt acctctgacg tttctatctc tccattgtat 2820
gttttttctt ggtttattct ttcatctctt ttctagcttc ttagtttagt catgccttct 2880
55 cctctaagtg ctgccttacc tgcacctgt gttttgatgt gaagtaact caacatcagc 2940
cactttcaag tgttcttaaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000
ctgtgatttt tttctttgtg cagctgtgtt tttagcgtga aatcattttg atatcagta 3060
cttttaagta ttcttttagt tattctgtga ttcttttag cagttagtta ttgaacact 3120
gttttatgtt aagatatgta gagtataaag atacgtagag tatttttaag tatcatttta 3180
60 ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240
ggagccttgc tttgtgatct agtgtgtgca tgggttccag aactgtccat tgtaaatttg 3300
acatccctgc aatagtgggc atgcatgttc actatatcca gcttattaag gtccagtgca 3360
aagcttctgt ctcttcttag atgcatgaaa ttccaagaag gaggccatag tccctcacct 3420
gggggatggg tctgttcatt tcttctcgtt tggtagcatt tatgtgaggc attgttaggt 3480
65 gcatgcacgt ggtagaattt ttatcttctc gatgagtga tcttttggag acttctatgt 3540

	ctctagtaat	ctagtaattc	tttttttaaa	ttgctcttag	tactgccaca	ctgggcttct	3600
	tttgattagt	attttctctg	tgtgtctggt	ttctgccttt	aatttatata	tatatatata	3660
	tttttttttt	ttttgagaca	gagtccttgg	ctgtcgccca	gggtgagtgc	agtggtgtga	3720
	tcacaggtca	gtgtaacttt	taccttctgg	cctgagccgt	cctctcacct	cagcctcctg	3780
5	agtagctgga	actgcagaca	cgcaccgcta	cacctggcta	atttttaaat	tttttctgga	3840
	gacagggctc	tgtgtgtgtg	cccaggtctg	tctcaaactc	ttggactcaa	gggatccatc	3900
	tacctcggct	tcccaaagtg	ctgaattaca	ggcatgagcc	accatgtctg	gcctaatttt	3960
	caacactttt	atattcttat	agtgtgggta	tgctctgtta	acagcatgta	gggtgaatttc	4020
10	caatccagtc	tgacagtcgt	tgtttaactg	gataacctga	tttattttca	tttttttgtc	4080
	actagagacc	cgcctgggtg	actctgattc	tccacttgcc	tgttgcatgt	cctcgttccc	4140
	ttgtttctca	ccacctcttg	gggtgccatg	tgcgtttctt	gccgagtgtg	tgttgatcct	4200
	ctcgttgect	cctggctcact	gggcatttgc	ttttatttct	ctttgcttag	tgttaccctc	4260
	tgatcttttt	attgtcgttg	tttgcctttg	tttattgaga	cagtctcact	ctgtcaccca	4320
	ggctggagtg	taatggcaca	atctcggctc	actgcaacct	ctgcctcctc	ggttcaagca	4380
15	gttctcattc	ctcaacctca	tgagttagctg	ggattacagg	cgcccaccac	cacgcctggc	4440
	taatttttgt	atttttagta	gagataggct	ttcaccatgt	tggccaggct	ggctctcaaa	4500
	tcctgacact	aagtgatctg	cccgcttggg	cctccacag	tgctgggatt	acaggtgcaa	4560
	gccaccgtgc	ccggcatacc	ttgatctttt	aaaatgaagt	ctgaaacatt	gctacccttg	4620
20	tcctgagcaa	taagaccctt	agtgtatttt	agctctggcc	acccccagc	ctgtgtgctg	4680
	ttttccctgc	tgacttagtt	ctatctcagg	catcttgaca	ccccacaaag	ctaagcatta	4740
	ttaatatgtt	tttccgtgtt	gagtgtttct	gtagctttgc	ccccgccctg	cttttctctc	4800
	tttgttcccc	gtctgtcttc	tgtctcaggc	ccgcctctct	gggtccccct	ccttgtctct	4860
	tgctgggttc	ttctgtcttg	ttattgcttg	taaaccocag	ctttacctgt	gctggcctcc	4920
25	atggcatcta	gcgacgtccg	gggacctctg	cttatgatgc	acagatgaag	atgtggagac	4980
	tcacagaggag	ggcgggtcatc	ttggcccgtg	agtgtctgga	gcaccacgtg	gccagcgttc	5040
	cttagccagt	gagtgcagac	aacgtccgct	cgccctgggt	tcagcctgga	aaacccacag	5100
	catgtcgggg	tctgggtggc	ccgcgggtgc	gagtttgaaa	tcgcgcaaac	ctgcgggtgtg	5160
	gcgccagctc	tgacgggtgt	gcctggcggtg	ggagtgtctg	cttcctccct	tctgcttggg	5220
	aaccaggaca	aaggatgagg	ctccgagccg	ttgtcgccca	acaggagcat	gacgtgagcc	5280
30	atgtggataa	ttttaaaatt	cttaggctgg	gcgcgggtgg	tcacgcctgt	aatcccagca	5340
	ctttgggagg	ccaaggcggg	tggatcacga	ggtcaggagg	tcgagaccat	cctggccaac	5400
	atgatgaaac	cccatctgta	ctaaaaacac	aaaaattagc	tgggcgtggg	ggcgggtgcc	5460
	tgtaatccca	gctactcggg	aggctgaggg	aggagaattg	cttgaacctg	ggagttggaa	5520
	gttgacagtga	gccgacattg	caccactgca	ctccagcctg	gcaacacagc	gagactctgt	5580
35	ctcaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aattctagta	gccacattaa	aaaagtaaaa	5640
	aagaaaagggt	gaaattaatg	taataataga	ttttactgaa	gccagcatg	tccacacctc	5700
	atcatttttag	gggtgtattg	gtgggagcat	cactcacagg	acatttgaca	ttttttgagc	5760
	tttgtctcgc	ggatcccggt	tgtaggtccc	gtgcgtggcc	atctcggcct	ggacctgctg	5820
40	ggcttcccat	ggccatggct	gttgtaccag	atgggtcagg	tccgggatga	ggtcgccagg	5880
	ccctcagtga	gctggatgtg	cagtgtccgg	atgggtgcacg	tctgggatga	ggtcgccagg	5940
	ccctgctgtg	agctggatgt	gtggtgtctg	gatggtgcag	gtcaggggtg	aggtctccag	6000
	gccctcgggt	agctggaggt	atggagtcgg	gatgatgcag	gtccgggggtg	aggtcgccag	6060
	ggcctgctgt	gagctggatg	tgtggtgtct	ggatgggtgca	ggtcaggggt	gaggtctcca	6120
45	ggccctcggg	aagctggagg	tatggagtcc	ggatgatgca	ggtcgggggt	gaggtcgcca	6180
	ggccctgctg	tgagctggat	gtgtggtgtc	tggatgggtg	aggtctgggg	tgaggtcacc	6240
	aggccctcgg	gtgagctggg	tgtgcgggtg	ctggatgggt	caggtctgga	gtgaggtcgc	6300
	cagacgggtg	cagaccatgc	ggtagctggg	atatgcgggt	tccggatggg	gcaggtctgg	6360
	ggtagggttg	ccaggccctg	ctgtgagttg	gatgtggggg	gtccggatgc	tgaggtcgg	6420
50	gtgtgaggtc	accaggccct	gctgtgagct	ggatgtgtgg	tgtctggatg	gtgcaggtct	6480
	gggggtgaagg	tcgccaggcc	cctgcttctg	agctggatgt	gtggtgtctg	gatggtgcag	6540
	gtctggagtg	aggtcgccag	gccctcgggt	agctggatgt	gcagtgtcca	gatggtgcag	6600
	gtccgggggtg	aggtcgccag	accctcgggt	gagctggatg	tcgggtgtct	ggatgggtgca	6660
	ggctctggagt	gaggtcgcca	ggccctcggg	gagctggatg	tatggagtcc	ggatgggtgca	6720
55	ggtcgggggt	gaggtcgcca	gaccctgctg	tgagctggat	gtgcgggtgtc	tggatgggtac	6780
	aggtctggag	tgaggtcgcc	agaccctgct	gtgagctgga	tatgcgggtgt	ccggatgggtg	6840
	caggtcaggg	gtgaggtctc	caggccctcg	gtgagctgga	ggtagtgagt	ccggatgggtg	6900
	caggtcgggg	gtgaggtcgc	caggccctgc	tgtgaactgg	atgtgcggcg	tctggatggg	6960
	gcaggtctgg	gggtgtggtc	ccaggccctc	ggtagctggg	aggtatggag	tccggatgat	7020
60	gcaggtccgg	ggtaggtcgc	ccaggccctg	ctgtgagctg	gatgtgcggc	gtctggatgg	7080
	tgacaggtctg	gggtgtggtc	gccaggccct	cgggtgagctg	gaggtatgga	gtccggatga	7140
	tgacaggtccg	gggtgaggtt	gccaggccct	gctgtgagct	ggatgtgctg	tatccggatg	7200
	gtgcagtcgg	gggtgaggtc	gccaggccct	gctgtgagct	ggatgtgctg	tatccggatg	7260
	gtgcaggtctc	gggtgaggtc	caccaggccc	tcgggtgagc	tgggtgtgctg	gtgtccgggt	7320
65	gtgcaggtctc	gggtgaggtc	caccaggccc	tcgggtgagc	tggatgtgctg	gtgtcccggt	7380
	gtccggatgg	tgacaggtcca	gggtgaggtc	gctaggccct	tgggtgggctg	gatgtgccgt	7440

gtccggatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgcggt 7500
 gtctgcatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgctgg 7560
 gtccggatgg tgcaggctccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgagg 7620
 5 tgtctggatg gtgcaggctcc ggggtgaggt agccaaggcc ttcggtagagc tggatgtggg 7680
 gtgtccggat ggtgcaggctc cggggtgagg tgcaggccctt ctgaggcttag ctggatatgc 7740
 ggtgtccgga tgggtgcaggc ccggggtgag gtcaccaggc cctgaggctta gctggatgtg 7800
 cgggtgtctgg atggtgcagg tccggggtga ggtcgccagg ccctgctgtg agctggatgt 7860
 gctgtatccg gatggtgcag gtccggggtg aggtcgccag gccctgcagt gagctggatg 7920
 10 tctgtatcc ggatggtgca ggtctggcgt gaggctcgcca ggccctgagg ttagctggat 7980
 atgagggtgc ggatggtgca ggtccggggt gaggctacca ggccctgagg ttagctggat 8040
 gtgagggtgc cggatggtgc aggtctgggg tgaggctgcc aggccctgct gtgagctgga 8100
 aagcgtgtat ccggatggtg caggctccgg gtgaggctgc caggccctgc ggtgagctga 8160
 atgtgtctga tccggatggt gcaggctctg cgtgaggctc ccaggccctg cggtagagctg 8220
 gatgtgcagt gtacggatgg tgcaggctcc gggtagaggtc gccaggccctt ggggtgggct 8280
 15 gtatgtgtgt tgtctggatg gtgcaggctc ggggtgagtt cggcaggccc tgggtgagc 8340
 tggatgtgtg gtgtctggat gctgcaggct cggggtgagt tgcaggccctt ctcggtagc 8400
 tggatatgag gtgtcccggt gtccgaatgg tgcaggctca gggtagaggtc gccaggccctt 8460
 tggtagagct gatgtgagg gtccggatgg tgcaggctct gggtagaggtc gccaggccctt 8520
 20 tggtagagct gatgtgagg gtccggatgg tgcaggctcc gggtagaggtc accaggccctt 8580
 cggtagatct gatgtggcat gtccttctcg ttttaag 8616

 <210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

 <400> 6
 gtactgtatc cccacgccag gcctctgctt ctggaagtc tggaaacacca gcccgccctc 60
 30 agcatgcgcc tgtctccact tgctctgctt tccctggctg tgcagctctg ggctgggagc 120
 caggggcccc gtccacaggcc tgggtccaagt ggattctgtg caaggctctg actgcctgga 180
 gctcacgttc tcttacttgt aaaatcagga gtttgtgcca agtggctctt aggggttgta 240
 aagcagaagg gatttaaatt agatggaac actaccacta gcctccttgc ctttccctgg 300
 gatgtgggtc tgattctctc tctctttttt ttttctttt tggatggag tctactctg 360
 35 ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
 ttaagcgatt caccagcctc agcctcctaa gttagctgga ttacaggcac ctgccaccac 480
 gcctggcctaa tttttgtact ttttaggagag acgggggttc accatgttgg ccaggctggt 540
 ctggaactca tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600
 ggctaagcca ccgtgccag ccccgatgc tcttttaatt catgctgttc tgatgaatc 660
 40 ttcaatctat tggatttagg tcatgagagg ataaaatccc accacttgg cgactcactg 720
 cagggagcac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780
 taggtggctg catttgaatg gctgtgagat tttgtctgca atgttcggct gatgagagt 840
 tgagattgtg acagattcaa gctggatttg catcagttag ggacgggagc gctggtctg 900
 gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960
 45 tgactgtgga gggctttagt cagaagatca gggcttcccc agctcccctg cacactcgag 1020
 tccctggggg gccttgtgac acccatgcc ccaaatcagg atgtctgcag agggagctgg 1080
 cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tgggtgctgg 1140
 gccatttctt tgcatctggg ggagggtcag ggccttccct gtgggaacaa gtttaatacac 1200
 aatgcacctt acttagactt tacacgtatt taatgggtg cgacccaaca tggatcattg 1260
 50 accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtggtggt 1320
 ccccaagatg ctcttgtgca ctactgggac tgtgtgtctg cctggggggc cttggaggcc 1380
 cctcctcccc ggacagggtg ccgtgccttt tctactctgc tgggcctgag gctgaggtc 1440
 agggcaccag ctccggagca ccccgggccc cagtgtccac ggagtgcag gctgtcagcc 1500
 acagatgccc aggtccagggt gtggcgctc cagccccctg gccccatgg gtggttttg 1560
 55 gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcatgagag ctgattctgc 1620
 tcttggctg agctgcctg agcagcctc cccgcccctc ccatctgaag ggatgtggct 1680
 ctttctacct ggggtcctg cctggggcca gccttgggt accccagtgg ctgtaccaga 1740
 gggacaggca tctgtgtggt aggggcatgg gttcacgtg cccagatgc agcctgggac 1800
 caggctcccc ggtgctgatg gtgggacagt caccctgggg gttgaccgccc ggactgggag 1860
 60 tccccagggg tgactatagg accaggtgtc cagggtgccct gcaagtagag gggctctcag 1920
 aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgtgccc 1980
 tgggtgcccc gagccctcac tgagtgggt ggggcttgtg gcttcccgtg agcttcccc 2040
 tagtctgttg tctggtgag caagcctcct gaggggctct ctattgcag 2089

10 / 18

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5

<400> 7
gtggctgtgc tttggtttaa ctctcttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccgggtgcc tccagaaaag 180
10 cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tcctgctgtg 300
gtgcgcagcc tccgtgctgt tccgtttacg gggcccgggg accaggccac gactgccagg 360
agccqaccgg gctctgagga tcctggacct tgccccacgg ctctctgcacc ccacctctgt 420
ggctgctggg gctgcgggtga ccccgctatc tggagagagt gtggggtgag gtggacagag 480
15 gtgtggcatg aggatcccg gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540
tctgaggaag ctgggagggg ttctaggtcc cgggtctggg tggctgggga cactggggag 600
gggctgcttc tcccctgggt ccctatggtg ggggtgggac ttggccggat ccactttcct 660
gactgtctcc catgctgtcc ccgccag 687

20

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

25

<400> 8
gtgggtgccc gggacccccg tgagcagccc tgctggacct tgggagtggc tgcttgattg 60
gcacctcatg ttgggtggag gaggtactcc tgggtggggc gcagggagtg caggtgacct 120
tgctactgtt gaggacacac ctggcaccta ggggtggaggc ctccagcctt tcctgcagca 180
catggggccc actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtcccactg 240
30 gattccagtt tccgtcagag aaggaaccgc aacggctcag ccaccaggcc ccggtgcctt 300
gcacccagct cctgagccag gggctcctg tcctgaggct cagagagggg acacagcccc 360
ccctgccctt ggggtctgga gtggtggggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag tttctgcgtg gccactgtca gtctcctcgc 480
ctccactcac acag 494

35

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

40

<400> 9
gtaaggttca cgtgtgatag tcgtgtccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgtgg cgtgcatatt 180
45 tgtggtgtgt gtgtgtgtgg cacgtgtgtg tccatgggtg gtgtgcctgt ggtgtgcatt 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgccta 300
tttgtggtgt gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360
ggcccccttg ccttactcct tctcctccca ggcattgtcc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggctcctc cttctagcat ggggtgcccc 480
50 gtctgtcac agggctgggc cttggagact gtaagccagg tttgagagga gagtagggat 540
gctggtggtg ccttctctga cccctggcac cccagggacc ccagtctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctcgctcccc gggacacact cctcccagag 660
cggccggggg ccttggggct cggcaggggt gaaagggggc ctgggcttgg gttcccaccc 720
agtgtcatg agcacgctgg aggggttaag cctcaaagtc gtgccaggcc ggggtgcaga 780
55 ggtgaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgacttctt gagct 865

60

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

<400> 10
tgtgggattg gttttcatgt gtgggatagg tgggatctg tgggattggt ttttatgagt 60
65 ggggtaaacac agagttcaag gcgagcttct tcctgtagt gggctctgag gtgctccaac 120

agctttattg aggagaccat atcttctctt gaactatggt cgggtttata gtaagtcagg 180
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tcgtgtgggt 240
 cctgctgtgg tgtgtggccg gtgggcaggg cttccaggcc tccttgtgtt cattggcctg 300
 gatgtggccc tggctacgct ccgtccttgg aattcccctg cgagtggag gctttctttc 360
 5 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgccc 420
 ggctggagtg gtttggcgtg atcttggctc actgcaacct gtgcttcctg agttcaagca 480
 attctcttgc ctcagcctcc caagtagctg gaattatagg cgtccaccac catgctgact 540
 aatttttcta atttttagtag agacgagggt tctccatgtt ggccaggctg gtctcgaact 600
 cctgacctca ggtgactctc ccacctcggc ctcccaaagt gctgggatga cagggtgtga 660
 10 ccgcccgcgc cggccgagac tcgcttctct cagcttccgt gagatctgca gcgatagctg 720
 cctgcagcct tgggtgctgac aacctccgtt tctcttctcc aggtctcgct aggggtcttt 780
 ccatttcag actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgtttcct 840
 gcgtaattgg tgtctgctgt ttatcgatgg cctccttcca tttcctttag gctttgttta 900
 ttgtgtttt tccggctcct tgaaggaaaa gtctcgatta tggatgtttg aactttcttt 960
 15 tctaaacaag catctgaagt tggcgttttc cctctaaagc agggatcccg agggccctgg 1020
 ctgtggagtg gcaccgggtc ggggcctgtt aggaaccggc cgcacagcgg gaggctaggt 1080
 ggggtgtggg gagccagcgt tcccgcctga gccccgcccc tctcagatca gcagtggcat 1140
 gcggtgtcca gagggcaca caccctactg agaactgtgc gtgagagggg tctagattct 1200
 gtgctcctta tgggaatcta atgctgtgag atctgaggtg gaaccgtttg ctcccaaac 1260
 20 catccccttc ccactgctg tctgtgggaa aaatcgctct ccacgaaacc agtccctggg 1320
 accacaatgg ttggggaccg tgtgtctaaag acctgcttca gcagcctctc gtcagtgttg 1380
 atatatggc ttttctgtgt tgagtcacaga ataattacgg atttctgtga tgctttccgc 1440
 cgacctcaga cccatgggct atttgtgggc gtgttgccct ctctgggtt gggaaggggtg 1500
 caggcccat gtaccttctt gttactgcct tccagggttg ttctcagggt tgaatcgta 1560
 25 tcgatgtgtt tttagccac ggccctgccc ccagctcctg ggggctgggg aacatgctga 1620
 agcacagagt caccgtgcgc gtcttttgat gcctcacaag ctcgaggcct cctgtgtccg 1680
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtcttggggg agtctgcaga ataggagggt ggggtgcccg 1800
 tctctctccc gcgtcttcag actcttctcc tgctgtgtgt gtggctgcac ctgcatccct 1860
 30 gcaatccctc cagcactggg ctggagaggc ccgggagctc gactgccact tgtgccact 1920
 gactgtggat ggcagtcggg caggggggtc tgatgtgtgt tgaactgtga tggcggttg 1980
 tcacaggggt ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga tgtgtgact 2040
 gtggatggcg gtcgtggggg ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga 2100
 tgtgtgtact tgggatggcg gtcgtggggg ctgatgtgtg gactgtggat gggggtctg 2160
 35 ggggtctgat tgggtactgt ggatggcagt cgtgggtct gatgtgtgt gactgtggat 2220
 ggcggctctg ggggtctgat tgggtactgt ggatggcagt cgtgggtct gatgtgtgt 2280
 gactgtggat ggcggctctg ggggtctgat tgtgtgtact gtggatggcg gtcgtgggg 2340
 ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga tgtgtgtga ctgtggatg 2400
 cggctctggg gtcgtgatgt gtgactgtgt atggcggtcg tggggtctga tgtgtgtga 2460
 40 ctgtggatgg tgatcggtca caggggtctg atgtgtgtgt actgtggatg gcggctctg 2520
 ggtctgatgt gtgtgtactg tggatgggtga tccgtcacag ggggtctgat tgtgtgtact 2580
 gtggatggcg gtcgtggggg ctgatgtgtg gtgactgtgt atggcggttg gtcccgggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtactgt 2700
 ggatggcggt cgtgggggtc gatgtgtgt gactgtggat ggcgggtcgt ggggtctgat 2760
 45 tgtgtgtact gtggatggcg gtcgtggggg ctgatgtgtt gactgtggat ggcgggtcgt 2820
 ggggtctgat tgggtactgt ggatggcggt cgtgggggtc gatgtgtgt gactgtggat 2880
 ggcgggtgtt cccgggggtc tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgat 2940
 gtgtgtactg tggatggcag tctgtgggtc tgatgtgtgt tgactgtgga tggcggtcgt 3000
 ggggtctgat gtgtgtgtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060
 50 gatggcggtc gtggggtctg atgtgtgtgt actgtggatg gcgggtcgtg ggtctgatgt 3120
 ggtgactgtg gatggcggtc gtggggtctg atgtgtgtgt actgtggatg gtgatcgtg 3180
 acaggggtct gatgtgtgtt gactgtggat ggcgggtcgt ggggtctgat tgtgtgtact 3240
 gtggatggcg gtcgtggggg ctgatgtgtt gactgtggat ggcgggtcgt ggggtctgat 3300
 tgtgtgtact gtggatggcg gtcgtagggt ctgatgtgt gtgactgtgt atggcagtcg 3360
 55 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtggggtctg atgtgtgtg 3420
 actgtggatg ggcgtcgtgg ggtctgatgt gtgggtactg tggatggcg tctgtgggtc 3480
 tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgat gtgggtactg tggatgggtga 3540
 tcggtcacag ggggtctgat tgtgttagct gcagggtggg tcccagggtg gtctgtagct 3600
 actttgcgtc ctggccccc cggcccccgt ttcccaaac gaagcttccc aggcgtctc 3660
 60 tgggcttcat ccgcccacgt ggcttggcgg caggtccaca cgtcctgac ggaagaaaca 3720
 agtggccagc tctggccggg gcaggccaca tttgtggctc atgcccctc ctctgccggc 3780
 ag 3782

12 / 18

<210> 11
 <211> 980
 <212> DNA
 <213> Homo sapiens

5

<400> 11
 gtctggggcac tgccttcgag gggtggggcac ggactcccag cagtgggtcc tcccctgggc 60
 aatcactggg ctcagaccg gacagactgt tggccctggg gggcagtggt gggaaatgagc 120
 tgtgatggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
 tgcgacagct gctgcattca ggcacctgct cacgtttgac tgcgaggcct ctctccagtt 240
 ccgagtgcc tttgttcag atttgctaaa tgtcttctct gccagttttg atcttgaggc 300
 caaaggaaag gtgtccccc cctttaggag ggcaggccat gtttgagccg tgtcttggcc 360
 agctggcccc tcagtgtctg gtctgagccc aaaggaaacg tgtcccccct cttaggagga 420
 cgggcccgtg ttgagccacg ccccgctgag cgggcccctc agtgctgggt ctgtccacgt 480
 ggccctgtgg ccttttgag atgtgtctg tccacgtggc cctgtggctc tttgcagatg 540
 cctgttagca cttgctcggc tctaggggac agtcgtgtcc accgcatgag gctcagagac 600
 ctctgggcga atttccttgg ctcccagggt ggggggtggg gtggcctggg ctgtgggac 660
 ccagaccctg tgcggggcag ctgggcagca actcctggat cacatatgcc atccgggcca 720
 cgggtgggctg tgtgggtgtg agcccagctg gaccacagg tggcccagag gagacgttct 780
 gtgtcacaca ctctgcctaa gcccattgtg gtctgcagag actcggcccg gccagcccac 840
 gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaggga 900
 cggaggggtct tggccacgtg gtccctgctg tctcagcacc caccggctca ctcccatgtg 960
 tctcccgtct gctttcgag 980

25

<210> 12
 <211> 2485
 <212> DNA
 <213> Homo sapiens

30

<400> 12
 gtgagtcagg tggccagggt ccattgcctt gcgggtgggt gggcgggctg gcagggtctc 60
 tgctcacctc tctcctgccc ctccccact gnccttctgc ccggggccac cagagtctcc 120
 tttcttgccc cccgccccct cgggctctg ggtgcaggc tcccgaggcc ccggaacat 180
 ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
 ggggtgtgga gttgtcctg cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300
 tgcgcccagc gtttgagcct gcagctgtc agctccaaat tactactgac gctggacacc 360
 cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420
 tcctgtccct gtctgtgtgac ccccgcgagg gcgcggtctc ttctctctgt gactagattt 480
 cccatctgga aagtgcgggg ttgaccgtgt agtttgcctc tctcgggggg cctgtggtg 540
 ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600
 cgggtggtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660
 cacacctccc ggcaggcatc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
 ggaggaaatt cgtgcacact caaggtcatc agcaaggcca tccgcagtcg ggtggaacgt 780
 ggaggcctct ctctgggagc gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840
 ttataataa atataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
 attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960
 cacaaattgc acatggcagc agagtgaatt ttggcggagg gacacgtgtg cacatgtgtg 1020
 taagcggccc ccaggccac agaattcgtc gacaaagtca cctcccaga gaagccacca 1080
 cgggcctcct tctgtgtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
 gtggcagggc tttggggaat gtgagtgat gactgcgtcc tcatgccctg acagacagga 1200
 ggtgactgtg tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260
 tggctccagt tggcctctga ataaaaacgt ctcaaaaacc tgtgcccga aaaaactaaga 1320
 acagagagag tttcccatcc catgtgtcca caggggcgta tctgcttgcg ttgactcgct 1380
 gggctggccg gactcctaga gttggtgcgt gtgcttctgt gcaaaaagtg cagtccctct 1440
 gcccatcact gtgatctctg caccagcaag gaaagcctct tttcttttct tctttttttt 1500
 ttttttgaga cggaaacgtc ctgtgtctg cctgggcttg agtgagtggt cgcatctca 1560
 actcactgca acctccgct cccgggttcc agcatttctc ctgcctcagc ctcccagca 1620
 gctgagatta caggcaccac cccctgcgc ctggctaatt tttgtatttt tagtagagag 1680
 gggttttttc catgttgccc aggtctgtct cgaactcctg acctcaggtg atccaccac 1740
 ctccgcccct caaagtgtgt ggattacagg tgtgagccat caccgccagc cggaaagcct 1800
 ctttttaagg tgaccaccta tagcgttccc cgaaaataac aggtcttgtt tttgcagtag 1860
 gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
 tcgctgggca gccatgcctt ctgtgtgac ctttaggttc caccgggcta tctgtctctc 1980
 actgtttgtc tgaaaacgca ccttggcat cctgttttgg agagtttctg cttctcgttg 2040
 gtcattgtga aactaggggc aaggttgtat ccgttggcgc gcagcggcta catgtagggg 2100

65

13 / 18

catgagtctt tcaccgtgga caaatcctt gaaaaaaaa aaaggagtcc ggtaagcat 2160
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220
 aaagaaaacc ttgatgattc agagcaagga tgtgggcaca cctgtggctg gatctgtttc 2280
 agccgcccc a gtgcatggtg agagtgggga gcagggttg tttgttcaga ggtctcatct 2340
 5 ggtagtgtttc tgaggtgttt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
 tctcacctgt gtcttccgc cccag 2485

<210> 13
 10 <211> 1984
 <212> DNA
 <213> Homo sapiens

<400> 13
 15 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60
 agtgttaata ttcttggtgc tctggagacc atgactgtc tgtcttgagg aaccagacaa 120
 ggttgacgcc ccttcttggc atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
 gggctccacg caggctctgt ccagggcca tgtccagagg cctcagggtc cagcaggcgg 240
 gagggccgct gccctgcatg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
 20 ctgtcacgtc acccaggttc cgtaggggtc cttggggaga tggggctggt gcagcctgag 360
 gcccacatc tcccagcagg ccctcgacag gtggcctgga ctgggcgctt cttcagccca 420
 ttgcccattc cacttgcatg ggttctacac ccaaggacgc acacacctaa atatcgtgcc 480
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540
 tactttaagt tctaggttac atgtgcacga cgtgcagggt agttacatat gtatacatgt 600
 25 gccatgttgg tgtgctgcac ccattaactc atcatttaca ttaggtatat ctccaatgc 660
 tatccctccc cactcccccc atcccatgac aggccctggt gtgtgatgtt cccaccctg 720
 tgtccaagt tctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttggtt 780
 ttctttcctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtggtgta tatgtgccac 900
 30 attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttctact 960
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
 tcctttgggt atataccag taatgggatg gctgggtcaa atggtatttc tagttctaga 1080
 tccttgagga atcaccacac tgtctccac aatggttgaa ctagtttaca ctcccacaa 1140
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
 35 tagtatcact gaacaagcag acagttagtg aaggatgctg caggaagcct gcaggccaca 1260
 cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320
 attatagcat gtacagtga tcaaggttct tcttcattaa ggttcaagtt ctgagattgaa 1380
 ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttga 1440
 ggaaagtgtc ctcgagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500
 40 ccatggtcat ggggcgctgg gcttggcct gagggtcaca cagtgcacca tgcccagctt 1560
 cctgtggata ggatctgggt ctcgatcat gctgaggacc acagctgcca tgctggtaaa 1620
 gggcaccacg tggctcagag gggcgaggt tcccagcccc agctttctta ccgtcttcag 1680
 ttatttttcc ctaagagtct gagaagtggg gccgcgctg atggccttcg ttcgtcttca 1740
 gctggcacag aattgcacaa gctgatggt aacactgagt acttataatg aatgaggaat 1800
 45 tgctgtagca gttactgta gagagctcgt ctgttgaaa gaaatttaag tttttcattt 1860
 aaccgctttg gagaatgtta cttattttat ggctgtgtaa attgtttgac attcagctcc 1920
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980
 ttag 1984

14 / 18

5 gggattgtcc aatgtggtcc ccctcaaggc cgccccacag agccggtggg cttgttttaa 660
 agtgcgattt gacgagggac gagaaacctt gaaagctgta aagggaaacc tcagaaaatg 720
 tggccgccag ggggtggttc aggtgctttg ctgggctgtg tttgtgaaaa ccattttgga 780
 cccgccctcc aagtccaccc tccaggtcca ccctccaggc ccgccctggg ctgggggtat 840
 gcctggcggt ccttgtgccc cagcccgagg cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cgggggggagc ccaggtccca agcaactgag ggctcaggag tcctgaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagttcc tgagggtgct 1020
 ggccagggag gtggctcaga gtgtatgttg ggggtcccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggc gtggccacag ggagctggga atgcaccagg 1140
 10 ggagctgcgc agctggccga ggtcccaggc ccaggccaca ggaagggcag ggggacgccc 1200
 ggggccacag cagaggccgc aggaaggga ggggatgcc aggccagagc agaggctacc 1260
 gggcacaggc ggggtccctg agctgggtga gcgaggctca tgactcggcg agggaacctc 1320
 cttgacgtga agctgacgac tgggtgtgcc cagctcacag ccagccagg tcccgcgcct 1380
 gagcaggaac tcagaacctt cccctttgtc taaagcacag cagatgcctt cagggcactt 1440
 15 aggagaaaac agggaaagtc gttgagaaac gtcttaaaag aagggtggat ggtggcaatt 1500
 tcttgtccag attttagtct gccccggacc acagatgagt ctataacggg atttgtgtgt 1560
 tgccatggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccactt 1620
 gagtctggc tgtcccggt ccaggccagg ttcttgcatg ctacctacc tgtctgccc 1680
 gggagacagg gaaagcaccc cgaagtctgg agcagggtg ggtccaggct cctcagagct 1740
 20 cctgccaggc ccagaccct gctccaaac accacttctc tggggtttcc caaagcattt 1800
 aacaagggtg tcaggttacc tcctgggtga cggccccga tcctggggct gacattgccc 1860
 ctctgcctta g 1871

25 <210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

30 <400> 15
 gtgagcgac ctggccgga gtggagcctg tgccccgctg gggcagggtg tgctgcagg 60
 ccgttgctc caectctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggc tgcccctcgt cccatctggg gctgagcaga aatgcattt tctgtgggag 180
 tgagggtgct cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240
 cagacctggg tgactgagg tgtcttcaga aagcagtctg gatccgaacc caagacgccc 300
 35 gggccctgct gggcgtgagt ctctcaaacc cgaacacagg ggcctgctg ggcagtgatc 360
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgccatgag ttcatgatca cgtgtgacct atcaggggac agggccatgg 540
 tgtggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccc agagctcaag 600
 40 gcccctctc aggtcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720
 ataatcccag cactttggga ggccgagggt ggtggatcac ttgaggccag gattttgagg 780
 ccaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 cctggtggca cagcctgta gtccccgcta tgccggaggc tgaggcagga gaatcattt 900
 45 aaccaggag gcagagggtg cagtgaaggc agatcacacc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
 ggacagggtg ttttttattc tgccttcga taatatttac tgggtgctgtg ctgaggccg 1080
 gaactggggg tgccttctc tgaaggcac accttcatgg gaagagaaat aagtgtgaa 1140
 tgggtgttaa accagagggt taaactgggg tcctgtcgtt ctgagttaac agtccagatc 1200

15 / 18

	gactggaagc	aaataagttg	tgtctttaca	gcatatacca	gagcagattc	taggtagaag	2220
	aggagacaca	tgcaaaacaac	accagcaaca	gaaataaaac	aaaagactca	aaggggaagg	2280
	aggtgaacgt	tccctggttt	ggtgttgagg	aaggacacac	agggaggcgg	atgaaaccag	2340
5	tgaggcaacg	ggcattgctt	tacttgacga	gaaactcagc	ttgcctgagc	cacagtgaag	2400
	atggccattc	cctggagcgt	ttgtgcacgt	gatttattta	aggcgccctg	tgaggtcctg	2460
	cacattcatc	ctctcacttt	gttctcctaa	ccacctgaga	ggtagaggag	gaaaggctcc	2520
	aggggagcag	cgcccttg	tcacccagct	ggcaaagggc	atgcatgatt	gcagcctggc	2580
	ctcctgctcc	ggggcccttg	ctctgcccga	ggaccccaca	caagtcagac	ccataggctc	2640
10	aggggtgagcc	ggagcccaag	gtcgtgttg	ggatggctgt	gaaagaagaa	atggacgtct	2700
	gatgcacact	tggaaggtc	ctaccagcag	cgtcaaagaa	atgcatgtga	aactgacagc	2760
	gagacccatc	cctcaaagaa	acgcacgtga	aactgatggc	gagacctgtc	cccatccctc	2820
	atgctggctc	cttttctggg	cttgccaaga	gccagcatca	ggttgaggca	agctggaaag	2880
	acttttctgg	aaagcagctt	gtttgcatgg	aagtcctcac	aatgtcctgt	gtcttcccag	2940
15	taattccact	tctgaagtga	ccagacatta	tcacgggtct	tatttaccat	ttccagtgtt	3000
	ccaggcaggg	ggacttgcca	cagcaagtca	cgaacctgcc	caaatacagg	gctaaggaga	3060
	tattatgcat	cacaaaactt	gctctgccat	taaacatttt	tcaaagaatt	tttgaagaat	3120
	gtttaatggc	acaaaacgtt	tatttcaatg	tagcagtgtt	caaagctgga	tgtaaaagaa	3180
	cacacccag	gagcctgccg	tgaatgtcat	gtgtgttcat	ctttggacat	ggacatacat	3240
20	gggcagtgag	tggtggtgag	gccctggagg	acatcggtgg	gatgcctcca	tcctgcccct	3300
	ctggagacac	catgtgtgcc	acgtgcactc	actggagccc	tgtttagctg	gtgccacctg	3360
	gctcttccat	ccctgagatt	caaacacagt	gagattcccc	acgcccact	cagtgttctc	3420
	ccacaaaaaa	cctgagtcac	acctgtgttc	actcgaggga	cgcccgggag	ccagggctcc	3480
	acagtttatt	atgtgttttt	ggctgagtta	tgtgcagatc	tcacagggc	agatgatgag	3540
25	tgcacaaaca	cgccgtgcg	aggtttgat	acactcaaca	tcactagcca	ggtcctggtg	3600
	gagtttggtc	atgcagagtc	tgatggcat	gtagcatttg	gagtcctggg	agtgagcacc	3660
	cagcccccctc	gggctgcagc	gcatgcccga	ggcaggacaa	ggaagcggga	ggaaggcagg	3720
	aggctctttg	gagcaagctt	tgacaggagg	ggctgggtgt	ggggcaggca	cctgtgtctg	3780
	acattcccc	ctgtgtctca	g				3801
30	<210> 16						
	<211> 880						
	<212> DNA						
	<213> Homo sapiens						
35	<400> 16						
	gtgagcaggc	tgatggctcag	cacagagttc	agagtccagg	aggtgtgtgc	gcaagtatgt	60
	gtgtgtgtgt	gtgcgcgcgt	gcctgcaagg	ctgatgggtga	ctggctgcac	gtaagagtgc	120
	acatgtacgc	atatacacgt	gagcacatac	atgtgtgcat	gtgtgtacat	gaaggcatgg	180
40	cagtgtgtgc	acaggtgtgc	aagggcacaa	gtgtgtgcac	atgcgaatgc	acacctgaca	240
	tgcatgtgtg	ttcgtgcaca	gtcgtgtggg	cattcacgtg	aggtgcatgc	gtgtgggtgt	300
	gcagtgtgag	tagcatgtgt	gcacataaca	tgtattgagg	ggtcctcgtg	ttcaccccgc	360
	taggtcctca	gcaccagtgc	cactccttac	aggatgagac	gggggtcccag	gccttggtgg	420
	gctgaggctc	tgaagctgca	gccttgaggg	cattgtccca	tctgggcac	cgctccact	480
45	ccctcctctg	tggtcttctg	tgtccactcc	ccctcctctg	tggtgattta	catccactcc	540
	actcctctc	tcctgtgggc	atccgcgtcc	actccccctc	tctgtgggca	tctgcgtcca	600
	cctccccctc	ctgtgggcat	ttgcgtccac	tccctcctct	ggttccttcc	tgtcttggtc	660
	gagcctcggg	ggcaggcaga	tgacacagag	tcttgactcg	cccagggtgg	ttcgcagctg	720
	ccgggtgagg	gccaggccgg	atttactctg	gaagagggat	agtcttctgt	caaaatgttc	780
50	ctctttcttg	ttccatctga	atggatgata	aagcaaaaaa	taaaaactta	aaatcccaga	840
	gaggtttcta	ccgtttctca	ctctttcttg	gcgactctag			880
	<210> 17						
	<211> 3186						
	<212> DNA						
55	<213> Homo sapiens						
	<400> 17						
	gtgagccgcc	accaaggggt	gcaggccag	cctccaggga	ccctccgcgc	tctgctcacc	60
	tctgaccggg	ggcttcacct	tggaactcct	gggttttagg	ggcaaggaat	gtcttacgtt	120
60	ttcagtggtg	ctgctgcctg	tgacagttc	tgctcgcgtg	gctctgtgca	aagcacctgt	180
	tctccatctc	tggttagtgg	taggagccgg	tggtggccca	ggtgtcccca	ctgtgcctgt	240
	gcactggccg	tggtgactga	tggtggccat	cccagggcag	caggggcatg	gggtaaagag	300
	atgtttatgg	ggagtcttag	cagaggaggg	tggtgaaggtg	tctgaacagt	agatgggaga	360
	tcagatgccc	ggaggatttg	gggtctcagc	aaagaggggc	gaggtgggtg	caggtgaggg	420
65	tcgttgcccc	cacccccggg	aaggtgcagc	agagctgtgg	ctccccacac	agccccggca	480

5 gcacactgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtccttg ctggtcaggg 540
 ggtgcccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
 acaggggcag cttctgcctg gactcagggc aggtggtggc acaagcctcg gggctgtacc 660
 aaagggcagt cgggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720
 ggagctgaat gccaggaggc cgaagccctc gcccctatgag ggctgagaag gactgtgagc 780
 atttgtgtta cccaggggcg aggtgcgcg aattaccgtg cacacttgat gtgaaatgag 840
 gtcgtcgtct atcgttgaaa cccagcaagg gctcacggga gacttttcca ttacaaggtc 900
 gtaccatgaa aatgggtttt aaccagagt gctgcgcctt catgctctgg cagggagggc 960
 10 agagccacag ctgcatgtta ccgcttttgc accagctcca gaggcttggg accaggctgt 1020
 ctcagttcca ggggtgcgtc ggctcagacc gccctcctct ctgcttcttc tctctgcctc 1080
 aaatcttccc tegtgtgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctcctttccg gaaacccttg ggggtgtgct gatacaggtg ccactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc agggctccagc tggcgtgctt ggggcctcct tgggcatga 1260
 tgaggtcaga ggagttttcc cagggtgaaa ctctgggaa actcccaggg ccatgtgacc 1320
 15 tgccacctgc tctctccata ttcagctcag tctgtcctc atttcccac cagggtctct 1380
 agctccgagg agctcccgtg gagggcctgg gctcaggga gggcggtga gtttcccac 1440
 ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgca 1500
 tggggccacgg gccgttttcca aacacagagt caggcacgtg gaaggccag gaatcccct 1560
 ccctcgaggc aggagtggga gaacggagag ctgggcccgc atttcacggc agccaggctg 1620
 20 cagtgggcca ggcgtgtggt gtccacgtgg cgctgggggc ggggtctgat tcaaatccgc 1680
 tggggctcgg ccttctctgg ccgtgctggc cgcgcctcca cacgggctg ggggtgacgc 1740
 cccgacctct agcagggtggc tatttctccc ttggaagag agcccctcac ccatgctagg 1800
 tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggcttattt 1860
 atttgtttta aaacattctg ggcctggctt ccgttgttgc taaatgggga aaagacatcc 1920
 25 cacctcagca gacttactga gaggtgaaa ccggggtgct ggcttgactg gtgtgatctc 1980
 aggtcattcc agaagtggct caggaaagta gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtgggt gagggccagg acatgggggg 2100
 ctcaggcact ggggtgagat aggtacacgg ggggctcagg cagagggtca gaccagggtac 2160
 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcattgtag 2220
 30 ccaggctctgt gcacacctgc cccaaagtcc caggaaagt agaggccaaa gatggaggct 2280
 gacagggctg gcgcggtggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 tatgaaaaaa aaaaacaaaa attagctgaa catggtgggt tgcgcctgta gttccaatac 2460
 35 ttgggaggct gaagtgggag gatcacttga gcccaggagg tggaaagtgc agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtgaag gcccatctca acaacaacaa 2580
 agaagactga caaatgcagt ttcttggaag gaaacattta gtaggaaact aacctacaca 2640
 cagaagccaa gtcggtgtct cggtgtcagt gagatgagat gatgggtcct cacaccatca 2700
 cccagacccc aggggtttatg caccacaggg gcgggtggct cagaagggat gcgcaggacg 2760
 ttgatatacg atgacatcaa ggtgtcttga cgaagggcag gattcatgat aagtacctgc 2820
 40 ttgtacacaa ggaacaaatg ataaactgga aaccttagag gccttcccgg aacaggggct 2880
 aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatgggtgca cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actcgacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060
 gcccattgag aaacccatgc atgtgcattc atgcacgcac acaggcaccg gtggggccat 3120
 45 gcccacaccc acgagcaccc tctgattagg aggcctttcc tctgacgctg tccgccatcc 3180
 tctcag 3186

<210> 18

<211> 781

<212> DNA

<213> Homo sapiens

<400> 18

55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
 ggagactgag tgaatctggg ctttaggaagt tcttaccctt ttctgcacaa ggaagtgggt 120
 taaccaaac actgtcaggc tegtctgccc gccctctcgt ggggtgagca gagcactga 180
 tggaggggac aggagctgtc tgggagctgc catccttccc accttgcctt gcctggggaa 240
 gcgctggggg gcctgggtctc tctgttttgc cccatgggtg gatttggggg gcctggcctc 300
 tctgttttgc cctgtgggtg gattgggctg tctcccgtcc atggcactta gggcccttct 360
 60 gcaaaccag gccaaaggct taggagagg ccaggcccag gctaccccac cctctcagg 420
 agcagaggcc gcgtatcacc acgacagagc cccgcgcctt cctctgcttc ccagtcaccg 480
 tctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aacctgcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

17 / 18

gtggaaattht cacctggaga agccgaagaa aacattttctg tctgtactcc tgcgggtgctt 660
 ggggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720
 gtgtctcctg ggaggggagc tgggtctggg ctgtgactcc tcagcctctg ttttccccc 780
 g 781

5
 <210> 19
 <211> 536
 <212> DNA
 <213> Homo sapiens

10
 <400> 19
 gcaagtgtgg gtggaggcca gtgcggggccc cacctgcccc ggggtcatcc ttgaacgccc 60
 tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cggccccggg cctgaccctg 120
 ggggccttga gccacgtctg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
 15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgatggctcc 240
 ccacgtctgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300
 ctgcccctgag ctctctgggt cctgagcaag ttctctcccc gccccgcgcg tccagcgtca 360
 ctgggctgcc tgtctgctcg ccccgggtga ggggtgtctg tcccttccact gaggttccca 420
 ccagccaggg ccacgaggtg caggccctgc ctgcccggcc acccacacgt cctaggaggg 480
 20 ttggaggatg ccacctctgg cctcttcttg aacggagtct gattttggcc ccgcag 536

<210> 20
 <211> 3179
 <212> DNA
 <213> Homo sapiens

25
 <400> 20
 atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
 ctgtgagtga acgggggtgt ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
 30 gtctatgagt gaatggggtt gtggtcagtg cgggcccatt gcctggctgg gcctgggagg 180
 tttctgatgc tgtgaggcag gagggggaag agggtagggg atagacagtg ggagccccc 240
 ccctggaaga cataacagta agtccaggcc cgaaggcgag cagggatgct gggggcccag 300
 ctggggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg gggggcccag 360
 ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
 35 ggctgggccc cctctcccc tgccctccac ctgcagccgt ggatccggat gtgcttccct 480
 ggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacacat 540
 cctgtataaa atccaggatt cctcctctg aacgcccaca ctcaggttga aagtcacatt 600
 ccgctcttgg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
 gtggggcagt ggagggtgtg gacacaggag gcttcagggt ggggtctggg atgctctctc 720
 40 atcctcttat catctccag tctcatctct catctctta tcctctccca gtctcatctg 780
 tcttctctct atctccagt ctcatctgtc atctctttac catctccag tctcatctct 840
 tatctcttta tctctagtct tcctccagac ttacctccca gggcggtgtc caggctcgca 900
 gtggagctgg acatacgtcc ttccctcaggc agaaggaact ggaaggattg cagagaacag 960
 gaggggcggc tcagagggac gcagtcttgg ggtgaagaaa cagccctccc tcagaagtgt 1020
 45 gcttgggcca cagaaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggtc 1080
 ccttggtggg ccttatggta tggccgggtc ctactgagtg caccttggac agggcttctg 1140
 gtttgagtgc agcccggacg tgcctggtgt cgggggtggg gcttatggcc actggatatg 1200
 ggcctattta ttgtgtctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260
 gggcccaagt ccacagactg tgcgtgaaat gcactctggt gcctggagcc cccgtatagg 1320
 50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaac 1380
 ggaaggaggc ggccccgggc gccgtgggag gacgacctca agtgagaggt tggacagaac 1440
 agggcgggga cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
 gaccaacagg tcaggccatt gttcagctat ccactctcta caaagctcca gattcctgtt 1560
 tctccgggtg ttttttgttg aaattttact caggattact tatatttttt gctaaagtat 1620
 55 tagaccctta aaaaagggtat ttgctttgat atggcttaac tcactaagca cctactttat 1680
 ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcacc 1740
 gggtgttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccccc aggtc 1800
 gatcctccgg cctcagcttc ccagagtgtg gggattacag gtgtgagcca ctgcccctgc 1860
 ctggcacttt taaaaaccac tatgtaaggt caggctcaggt ggcttccaca cctgtcatcc 1920
 60 cagtatgttg ggaagccgag gcagaaggat tgtctgaggc caggagtgtg agaccagcat 1980
 gggtaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtgggg 2040
 ccagcatctg tagtcccagc tgcctgggag gctgagtggg aggatcgctt gagcccgga 2100
 ggtcatggct gcagtgagct gtgattgtac catcgactc cagcctgggc aacagagtga 2160
 gaccctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
 65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaaggaa gaaggaggcc tgctagggtg 2280

5 taggtagact gtcaaatctc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340
aaccacagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttcacac 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaaagg gaggagaagc aggcaagggg 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520
gccccacgct cctgccggtc ctgcacctgc tgaaccgtc gatgttggtg ccagggtgcc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttgtgg caggcacaat tacagccctt ccccaaagat gccacgtcc ttctctctga 2700
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcaggtg gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatcttg tgggcctgat atggccacaa 2820
10 gggtccttag aagtgagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
15 cagtacaggg aaatgaatac agggacagtt ctacagtgga ctctcagccc acccctggg 3179